DOCUMENT RESUME

ED 038 061 48 FL 001 665

AUTHOR	Smith, Philip D., Jr.
TITLE	A Comparison Study of the Effectiveness of the Traditional and Audiolingual Approaches to Foreign Language Instruction Utilizing Laboratory Equipment. Supplementary Report.
INSTITUTION	West Chester State Coll., Pa.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.; Office of Education (DHEW), Washington, D.C. Inst. of International Studies.
BUREAU NO	BR-7-0133
PUB DATE	Aug 69
CONTRACT	OEC-1-7-070133-0445
NOTE	171p.
EDRS PRICE	EDRS Price MF-$0.75 HC-$8.65
DESCRIPTORS	*Audiolingual Methods, Comparative Analysis, Educational Research, Educational Strategies, Evaluation, French, German, *Grammar Translation Method, *Language Instruction, *Language Laboratories, Language Research, Modern Languages, National Norms, *Secondary Schools, Second Language Learning, Statistical Data, Testing, Textbook Evaluation
IDENTIFIERS	*Pennsylvania Foreign Language Project

ABSTRACT

USOE Projects 5-0683 and 7-0133 (Pennsylvania Foreign Language Projects entitled "An Assessment of Three Foreign Language Teaching Strategies Utilizing Three Language Laboratory Systems" and "A Comparison Study of the Effectiveness of the Traditional and Audiolingual Approaches to Foreign Language Instruction Utilizing Laboratory Equipment") were continued into Levels III and IV. The conclusions support earlier findings that students learning foreign language through a cognitive approach continue to achieve significantly better than audiolingual students. This supplementary report contains four major sections: (1) a review of the first and second years of the study, (2) a report of the third and fourth years, (3) additional information and analyses of the first and second years, and (4) reactions and reviews of the research discussion conference on USOE projects 5-0683 and 7-0133. The appendixes include a talk given by Dr. Smith in reply to discussion of the project which appeared in the October 1969 "Modern Language Journal". Extensive use of tables is made throughout the report. For related documents see ED 021 512 and ED 030 013. (Author/RL)

ED038061

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

PA-48
Title VI
BR 7-0133
OE/BR IIS
FL

SUPPLEMENTARY REPORT

PROJECT NO. 7-0133

CONTRACT
[redacted] NO. OEC-1-7-070133-0445

A COMPARISON STUDY OF THE EFFECTIVENESS OF THE TRADITIONAL AND AUDIOLINGUAL APPROACHES TO FOREIGN LANGUAGE INSTRUCTION UTILIZING LABORATORY EQUIPMENT

August, 1969

U.S. DEPARTMENT OF HEALTH,
EDUCATION, AND WELFARE
Office of education
Institute of International Studies.

FL 001 665

ED038061

BR 7-0133

SUPPLEMENTARY REPORT
Project No. 7-0133
Grant No. OEC-1-7-070133-0445
Title VI - N.D.E.A.

A COMPARISON STUDY OF THE EFFECTIVENESS OF THE TRADITIONAL AND AUDIOLINGUAL APPROACHES TO FOREIGN LANGUAGE INSTRUCTION UTILIZING LABORATORY EQUIPMENT

Author

Philip D. Smith, Jr.

Center for Foreign Language Research and Services

West Chester State College
West Chester, Pennsylvania

September 1969

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Institute of International Studies

COOPERATIVE RESEARCH PROJECT

Consultants

Dr. Robert Lado	Georgetown University
Dr. Stanley Sapon	University of Rochester
Dr. Wilmarth H. Starr	New York University
Dr. W. Freeman Twaddell	Brown University
Dr. Albert Valdman	Indiana University
Dr. Donald D. Walsh	MLA Foreign Language Program
Dr. Chauncy Dayton	University of Maryland
Dr. Rebecca Valette	Boston College

COOPERATING AGENCIES

United States Office of Education

Pennsylvania Department of Education

West Chester State College

Office of the Superintendents of Twenty Counties

Pennsylvania Public School Districts

PROJECT STAFF

Mr. Emanuel Berger	Principal Investigator
Dr. Philip D. Smith, Jr.	Coordinator
Dr. Alfred D. Roberts	Supervisor of Instruction
Mr. Ralph A. Eisenstadt	Field Consultant
Mr. William E. McDonald	Field Consultant
Mrs. Mary Ellen Allen	Field Consultant
Mrs. Constance Karr	Project Secretary

FOREWORD

On June 17, 1969, I was greatly pleased to be able to accept on behalf of the Pennsylvania Foreign Language Project the Governor's Award for Excellence for service to the Commonwealth of Pennsylvania. The award was made by Governor Raymond P. Shafer in a ceremony in the state capitol.

In presenting the award, Governor Shafer said, "This project... is the most extensive survey of the methods used in teaching foreign languages in secondary schools ever conducted in any state in the union." It was our understanding that this honor is not often awarded to educators.

The presentation ceremony was made even more meaningful when immediately afterward Ed DiMaio of George Washington High School, Project Student 21307, a member of the experimental population from Level I through Level IV, was honored by the Governor as Pennsylvania's "Teenager of the Year."

The Governor's Award is a direct reflection of the splendid cooperation of participating schools, teachers, administrators and students; the invaluable assistance of my friends of the Project Staff; and the constant support and guidance of Emanuel Berger and Robert B. Hayes of the Bureau of Research, the Pennsylvania Department of Education. The Governor's Award is ours, not mine. I merely have the custody for the entire Project team.

Philip D. Smith
West Chester State College
July 21, 1969

TO THE READER

The SUPPLEMENTARY REPORT OF USOE PROJECT 7-0133 has been produced to provide additional information on the effects of various teaching strategies on student achievement in foreign languages at the secondary school level. As its title states, it is to _supplement_ the previously completed reports, to correct errors of omission and reproduction, provide additional data analyses, to answer questions not previously treated, to extend and to discuss.

For this reason the document has little unity within itself. It can only relate to the Final Reports of USOE Projects 5-0683 and 7-0133. The SUPPLEMENTARY REPORT should not be read and studied alone.

AVAILABILITY OF DATA

Complete student data for the Pennsylvania Foreign Language Research Project are available to interested professionals. This information will be duplicated upon receipt of a blank standard 600 foot reel of one-half inch computer tape (800 b.p.i., 9 track IBM 360 system.)

Requests should be addressed to:

Director, Computer Center
Learning Research Center
West Chester State College
West Chester, Pennsylvania 19380

SCOPE OF THE PENNSYLVANIA FOREIGN LANGUAGE RESEARCH PROJECTS:

AN ASSESSMENT OF THREE FOREIGN LANGUAGE TEACHING STRATEGIES UTILIZING THREE LANGUAGE LABORATORY SYSTEMS

Title VII-A NDEA, Project 5-0683	$161,198.00

A COMPARISON STUDY OF THE EFFECTIVENESS OF THE TRADITIONAL AND AUDIOLINGUAL APPROACHES TO FOREIGN LANGUAGE INSTRUCTION UTILIZING LABORATORY EQUIPMENT

Title VI - NDEA, Project 7-0133	68,590.00

A Research Project of the Bureau of Research of the Department of Education with field headquarters at the Cooperative Research Center, West Chester State College.

Department of Education	22,098.00
Participating School Districts	45,857.00
TOTAL COMMITMENT	$ 297,743.00

Additional funds from Title V NDEA were utilized for test development and considerable direct support furnished by West Chester State College.

The Project was activated May 1, 1965 and extended through November 30, 1969.

Summary of Involvement

Original population:		3,500 students
First year completing:	104 classes;	2,171 students
Second year completing:	51 classes;	1,090 students
Replication:	28 classes;	639 students
Third year completing:	24 classes	
Fourth year completing:	17 classes	

TABLE OF CONTENTS

page

- FOREWORD i
- TO THE READER ii
- SCOPE OF THE PENNSYLVANIA FOREIGN LANGUAGE RESEARCH PROJECTS iii
- LIST OF TABLES vi
- LIST OF FIGURES viii
- SUMMARY ix
- SECTION I, REVIEW OF THE FIRST AND SECOND YEARS OF THE STUDY 1
  - INSTRUMENTATION 5
  - DEFINITIONS OF TEACHING STRATEGIES 6
  - LANGUAGE LABORATORY SYSTEMS 8
  - SUMMARY OF INSTRUCTIONAL CONTROLS 9
  - ORIGINAL HYPOTHESES 9
  - ANALYSES OF DATA 12
  - FIRST YEAR CONCLUSIONS 12
  - SECOND YEAR CONTINUATION AND REPLICATION 15
  - FIRST AND SECOND YEAR DISCUSSIONS AND IMPLICATIONS . . . 16
  - FIRST AND SECOND YEAR RECOMMENDATIONS 18
- SECTION II, THIRD AND FOURTH YEARS OF THE STUDY 20
  - INTRODUCTION 20
  - THIRD YEAR OBJECTIVES 22
  - THIRD YEAR FOLLOW-UP 22
  - THIRD YEAR EVALUATION 23
  - RESULTS OF THE COMPARISON OF TEACHING STRATEGIES, THIRD YEAR 23
  - ANALYSIS OF COVARIANCE, GERMAN III 25
  - INFLUENCE OF PRIOR LANGUAGE LABORATORY EXPERIENCE . . . 31
  - PREDICTION OF ACHIEVEMENT, GERMAN III 35
  - TEACHER PROFICIENCY AND STUDENT ACHIEVEMENT 35
  - STUDENT OPINION CHANGES 40
  - THIRD YEAR SUMMARY 41
  - FOURTH YEAR OF OBSERVATION 41
  - ANALYSIS OF COVARIANCE 41
  - REGRESSION ANALYSIS 43
  - SUMMARY 44
  - FOURTH YEAR STUDENT VIEWS 44
  - SUBSEQUENT IMPACT OF THE RESEARCH ON CURRICULUM AND LANGUAGE LABORATORY USAGE PATTERNS 49
- SECTION III, ADDITIONAL INFORMATION AND ANALYSES OF THE FIRST AND SECOND YEARS OF THE STUDY 53
  - DEFINITIONS OF STRATEGIES 53
  - TEACHER ABILITY AND PREPARATION 54
  - STATISTICAL COMPARISON 54
  - PRE-EXPERIMENTAL WORKSHOP 57
  - SUBSEQUENT PROFESSIONAL STATUS OF TEACHERS 59
  - INFLUENCE OF TEACHER NDEA INSTITUTE TRAINING ON CLASS ACHIEVEMENT 60
  - MLA COOPERATIVE CLASSROOM TESTS 60
  - STUDENT GRADE PLACEMENT 62

TABLE OF CONTENTS
(Continued)

page
ADDITIONAL ANALYSIS OF LEVELS I AND II 63
TEST SCORER RELIABILITY 68
ANALYSIS OF COVARIANCE BY TEACHING STRATEGY 69
ANALYSIS OF COVARIANCE BY LABORATORY SYSTEM 73
LANGUAGE LABORATORY USAGE 74
CONCLUSIONS 75
STUDENT OPINION SHIFTS 77
REGRESSION ANALYSES 80
SECTION IV, REACTIONS AND REVIEWS OF THE RESEARCH 84
DISCUSSION CONFERENCE ON USOE PROJECTS 5-0683 and 7-0133, WEST CHESTER, PA., MARCH 20, 1969 85
LUNCH DISCUSSIONS 101
AFTERNOON SESSION 102
Evaluation of PDE/FL Research Projects - Valdman 123
Some Conclusions to be Drawn from the Pennsylvania Study - Valette 127
Exerpts from the minutes of the February 1969 meeting of COFLIC 130
Memorandum to COFLIC members - 2/13/69 - Lester 131
Letter to State FL Supervisor - Lester 132
Letter to Lester, Foreign Language Consultant - Smith 133

APPENDIXES

A. Summary of Statistical Analyses, First and Second Years of the Study
B. Third and Fourth Year Classes and Schools
C. Project Norms, MLA Cooperative Classroom Test (MA)
D. List of Manufacturers of Language Laboratory Equipment for Schools and Treatments
E. In Reply to the October 1969 Modern Language Journal

LIST OF TABLES

page

TABLE 1 ANALYSIS OF VARIANCE BY STRATEGY, GERMAN III: FINAL MA LISTENING TEST 24
TABLE 2 ANALYSIS OF VARIANCE BY STRATEGY, GERMAN III: FINAL MA READING TEST 25
TABLE 3 ANALYSIS OF COVARIANCE BY STRATEGY - GERMAN III (Listening) 27
TABLE 4 ANALYSIS OF COVARIANCE BY STRATEGY - GERMAN III (Listening) 28
TABLE 5 ANALYSIS OF COVARIANCE BY STRATEGY - GERMAN III (Reading) 29
TABLE 6 ANALYSIS OF COVARIANCE BY STRATEGY - GERMAN III (Reading) 30
TABLE 7 ANALYSIS OF VARIANCE BY SYSTEM, FRENCH III: FINAL MA LISTENING TEST 31
TABLE 8 ANALYSIS OF VARIANCE BY SYSTEM, FRENCH III: FINAL MA READING TEST 32
TABLE 9 ANALYSIS OF VARIANCE BY SYSTEM, GERMAN III: FINAL MA LISTENING TEST 33
TABLE 10 ANALYSIS OF VARIANCE BY SYSTEM, GERMAN III: FINAL MA READING TEST 34
TABLE 11 MEANS AND CORRELATIONS, PRE-EXPERIMENTAL VARIABLES AND SUBSEQUENT GERMAN III ACHIEVEMENT . . 36
TABLE 12 MULTIPLE REGRESSION EQUATIONS FINAL GERMAN III ACHIEVEMENT 37, 38
TABLE 13 TEACHER, STUDENT DATA AND CORRELATIONS BETWEEN TEACHER PROFICIENCY SCORES AND CLASS ACHIEVEMENT AFTER THREE YEARS 39
TABLE 14 GERMAN STUDENT OPINION CHANGES, THREE YEAR PERIOD 40
TABLE 15 ANALYSIS OF COVARIANCE BY STRATEGY - GERMAN IV (Listening 42
TABLE 16 ANALYSIS OF COVARIANCE BY STRATEGY - GERMAN IV (Reading) 43
TABLE 17 REGRESSION ANALYSES, LEVEL IV CORRELATIONS . . 45
TABLE 18 STUDENT VIEWS ON FOREIGN LANGUAGE STUDY 46, 47
TABLE 19 COMPARISON OF TEACHER EXPERIENCE FACTORS . . . 55
TABLE 20 COMPARISON OF TEACHER PROFICIENCY FACTORS BY STRATEGY, FRENCH 56
TABLE 21 COMPARISON OF TEACHER PROFICIENCY FACTORS BY STRATEGY, GERMAN 57
TABLE 22 INFLUENCE OF TEACHER N.D.E.A. INSTITUTE TRAINING ON LATER CLASS ACHIEVEMENT 61
TABLE 23 ANALYSIS OF COVARIANCE BY STRATEGY, FRENCH I . 64, 65
TABLE 24 ANALYSIS OF COVARIANCE BY STRATEGY, GERMAN I . 66, 67
TABLE 25 SCORER RELIABILITY, SPEAKING TEST 68
TABLE 26 ANALYSIS OF COVARIANCE BY TEACHING STRATEGY FRENCH II 69, 70
TABLE 27 ANALYSIS OF COVARIANCE BY TEACHING STRATEGY GERMAN II 71, 72

LIST OF TABLES
(Continued)

TABLE 28 ANALYSIS OF COVARIANCE BY LABORATORY SYSTEM 73, 74
TABLE 29 SUMMARY OF REANALYSES 76
TABLE 30 ANALYSIS OF COVARIANCE OF FRENCH I STUDENT OPINION SHIFTS 77, 78
TABLE 31 ANALYSIS OF COVARIANCE OF FRENCH II STUDENT OPINION SHIFTS 78
TABLE 32 ANALYSIS OF COVARIANCE OF GERMAN I STUDENT OPINION SHIFTS 79
TABLE 33 ANALYSIS OF COVARIANCE OF GERMAN II STUDENT OPINION SHIFTS 79, 80
TABLE 34 REGRESSION ANALYSIS: FRENCH I FINAL ACHIEVEMENT 81
TABLE 35 REGRESSION ANALYSIS: GERMAN I FINAL ACHIEVEment 82
TABLE 36 REGRESSION ANALYSIS, STUDENT AND TEACHER EFFECTS ON MEAN CLASS ACHIEVEMENT, LEVEL I 83

LIST OF FIGURES

page
FIGURE 1 DISTRIBUTION OF CLASSES BY TEACHING STRATEGY AND LABORATORY SYSTEM 4
FIGURE 2 DISTRIBUTION OF GERMAN III STUDENTS 23

SUMMARY

USOE Projects 5-0683, AN ASSESSMENT OF THREE FOREIGN LANGUAGE TEACHING STRATEGIES UTILIZING THREE LANGUAGE LABORATORY SYSTEMS, and 7-0133, A COMPARISON STUDY OF THE EFFECTIVENESS OF THE TRADITIONAL AND AUDIOLINGUAL APPROACHES TO FOREIGN LANGUAGE INSTRUCTION UTILIZING LANGUAGE LABORATORY EQUIPMENT, were conducted by the Pennsylvania State Department of Education and West Chester State College during the 1965-66 and 1966-67 school years.

In essence, these studies failed to demonstrate any significant differences between "Traditional" and "Audiolingual" approaches on the *MLA Cooperative Classroom Listening* and *Speaking Tests*. Some significant differences existed in favor of the "Traditional" strategy on the *MLA Reading* and *Writing Tests*. The Language laboratory, regardless of type used twice weekly, had no discernable effect on class achievement.

With the encouragement of the Institute for International Studies of the U.S. Office of Education, the study was extended to permit observation of students through their third and fourth years of foreign language instruction. "Traditional" students continued to equal or significantly exceed "Audiolingual" students.

Reactions and criticisms of the study were solicited for inclusion in the SUPPLEMENTARY REPORT. These revealed a number of unanswered or unclear areas in reporting the first and second years of the study for which additional data and analyses are provided for the reader.

SECTION I

REVIEW OF THE FIRST AND SECOND YEARS OF THE STUDY

A Synopsis of

the Final Reports of USOE Projects 5-0683 and 7-0133

A COMPARISON STUDY OF THE EFFECTIVENESS OF THE TRADITIONAL AND AUDIOLINGUAL APPROACHES TO FOREIGN LANGUAGE INSTRUCTION UTILIZING LABORATORY EQUIPMENT

Pennsylvania has long been committed to a leadership role in the teaching of foreign languages. Testimony to this commitment is illustrated in the hundreds of language laboratories installed in its schools and the hundreds of teachers who have attended NDEA Foreign Language Institutes. Enrollment in foreign language courses includes one-third of the secondary school population. By 1965, every Pennsylvania public secondary school included foreign language instruction in the curriculum. In support of the foreign language program the State has mandated that, "... a minimum of a four-year sequence of a modern foreign language shall be offered by each school system" and required for certification that candidates receive passable scores on the skills portions of the _MLA Foreign Language Proficiency Test for Teachers and Advanced Students_.

Implicit in this strong state support for the teaching of languages is the responsibility to provide advice on problems of teaching methodology. It was therefore important that the profession initiate a study for investigating several basic unanswered problems related to secondary school foreign language instruction: (1) given several alternative teaching approaches to foreign language instruction which of these is better?, (2) which of the commonly used language laboratory systems is most effective as an adjunct to foreign language instruction? and (3) to study the relationship of the _MLA Foreign Language Proficiency Tests for Teachers and Advanced Students_ to student achievement.

Although this research was conducted in Pennsylvania, the results may be applicable to many schools throughout the nation. This was attempted by utilizing a large number of socio-economically representative schools and by minimizing the degree to which typical teaching conditions were to be modified. The instructional and testing materials were those commonly used in the teaching of foreign languages in the secondary schools.

Traditionally, foreign language instruction stresses student mastery of the formal grammar of the target language. The textbook, consisting of carefully graded reading selections and accompanying grammar lessons, is the essential pedagogical tool. The assumption is that proficiency in the language can be acquired by learning a set of grammatical rules to which the language is supposed to conform and by consciously applying these rules.

The "audiolingual" emphasis in modern foreign language teaching has roots extending back many years and is in sharp contrast to the formalistic traditional teaching methods. Many linguists believe that language learning is largely a behavioral skill and not an intellectual discipline. Developing this skill, like any other, requires the careful cultivation of language habits that are an automatic, almost unconscious performance of highly complicated physical and mental processes. In place of sole reliance on the textbook, the audiolingual teacher employs a set of teaching techniques and material specifically designed to develop oral and listening facility. The "dialogue" rather than the reading selection is the primary instructional tool for the beginning student.

This emphasis on imitation, practice, and repetition to the point of "over-learning" encouraged many schools that adopted the audiolingual approach to install language laboratory facilities. In the laboratory, each student is able to practice individually without disturbing other students. In addition, Hayes (1963) notes that the language laboratory provides native models of the foreign language for imitation, extensive structure drills, a variety of native voices necessary for understanding the language in its natural setting, and facilities for testing each student for listening and speaking ability.

In surveying the enormous research literature of foreign language teaching, most of the efforts following the broadly comparative Agard-Dunkel (1948) study have consisted of materials developed for audiolingual instruction and little useful research comparing new and conventional programs was possible (Birkmaier, 1960). Carroll (1963) dismisses most of the available studies as being "poorly controlled or otherwise deficient from the standpoint of valid research methodology."

Until 1965, no sufficiently realistic and generalizable research had been undertaken to shed light on which strategy or laboratory system works best when translated from a specific local small scale setting into the larger reality of numerous secondary schools. To assist in developing answers to this question, Pennsylvania undertook the large-scale in situ experiment which has come to be known as "The Pennsylvania Foreign Language Study." The research, a cooperative effort of the Bureau of Research, Department of Education and West Chester State College, was supported by grants under Titles VI and VII of the National Defense Education Act by the United States Office of Education.

A select group of foreign language educators was empaneled to help develop precise definitions of the distinguishing characteristics of each instructional strategy and to identify representative teaching materials.* These criteria are reproduced in a later portion of the SUPPLEMENTARY REPORT. A competent research staff was assembled and experimental guidelines were developed in great detail.

One hundred and four French I and German I teachers were identified who were willing to limit themselves to the experimental framework. Schools were located throughout the state and were judged to be a good representation of the secondary schools of the Commonwealth. Schools selected represented both "inner city" and suburban Philadelphia and Pittsburgh as well as a large number of diversified small communities. Students were from grades 8-12 with the majority in grades 9 and 10.

"Traditional" classes were taught, in the main, by teachers who preferred that strategy. It was possible to completely randomly assign eighty-seven classes among the "Audiolingual" and a modified "Audiolingual with Grammar" strategies. In addition, fifty-three classes could be randomly assigned to either the listen-respond or the listen-respond-record language laboratory system. A complete illustration of the assignment of experimental treatments is shown in Figure 1. In the final statistical analyses, only classes truly randomly assigned to laboratory treatment were considered.

Teachers were tested for foreign language proficiency and professional background with the state required MLA Teacher Proficiency battery and trained in their role at a week-long pre-experimental workshop. Three other meetings during the year facilitated research staff-teacher communication. The research staff observed teachers throughout the year on an unannounced irregular basis to insure adherence to strategy. Teachers averaged 9.9 years experience and forty-five graduate hours of preparation. Recent college graduates or residents abroad were excluded. Forty per cent of the teachers--twice the state average--had participated in National Defense Education Act Institutes and sixty-two per cent had traveled or studied abroad.

---

*Robert Lado, Dean, Institute of Languages and Linguistics, Georgetown University
Stanley Sapon, Dept. of Linguistics, University of Rochester
Wilmarth Starr, Dept. of German, New York University
W. Freeman Twaddell, Dept. of German, Brown University
Albert Valdman, Dept. of Linguistics, Indiana University, and
Donald D. Walsh, Foreign Language Program, Modern Language Asso. of America

FIGURE 1

DISTRIBUTION OF CLASSES BY TEACHING STRATEGY AND LABORATORY SYSTEM

Original Classes

French I, 1965-66

TLM	11

	TR	AA	AR*
AL+G	3	12	8
ALM	3	17	7

German I, 1965-66

TLM	6

	TR	AA	AR*
AL+G	5	9	4
ALM	4	10	9

Continuing Classes

French II, 1966-67

TLM	4

	TR	AA*	AR*
AL+G	2	5	7
ALM	2	3	2

German II, 1966-67

TLM	6

	TR	AA*	AR*
AL+G	4	2	3
ALM	3	6	2

Replicating Classes

French I, 1966-67

TLM	3

	AA*	AR*
AL+G	8	8
ALM	7	7

German I, 1966-67

TLM	1

	AA*	AR*
AL+G	5	5
ALM	4	4

*Classes split randomly into AA and AR sub-classes

Key:
TLM = Traditional
ALM = Audiolingual
AL+G = Audiolingual + Grammar
TR = Tape recorder only
AA = Additional Audio-Active (Listen-Respond) language laboratory drill
AR = Additional Audio-Record (Listen-Respond) Record) language laboratory

Representative texts for both approaches selected by the panel of foreign language specialists were those most widely used in the field.* Tests were of both the "new" and the "old" philosophy--The Modern Language Association *Cooperative Classroom* series and the 1939-41 *Cooperative French/German Tests.*

Each teacher in the audiolingual strategies used a tape recorder daily in the classroom. Classes assigned to one of the two laboratory periods spent two additional half-periods a week in laboratory practice with the commercially prepared tape programs. While decried by many foreign language educators as inadequate, the twice weekly laboratory practice was determined by surveys to be representative of existing administrative practice both before and after the experiment.

INSTRUMENTATION

Foreign Language Behavior

1. Listening Discrimination	*Valette Listening Discrimination Test*
	MLA Cooperative Classroom Test, 1963
2. Listening Comprehension	a. Listening
3. Speaking	b. Speaking
4. Writing	c. Writing
	Cooperative French (German) Tests, 1939-41
5. Reading	a. Reading
6. Grammar	b. Grammar
7. Vocabulary	c. Vocabulary
8. Expectations	*Student Expectations Scale*
9. Attitudes	*Student Opinion Scale* (semantic differential)

---

*Traditional: French, *Cours Elementaire de Francais* (1st and 2nd ed.,) 1949, 1956 *New First Year French*, 1958
German, *A First Course in German* (2nd ed.,) 1964 *Foundation Course in German* (Rev. ed.,) 1964
Audiolingual: French, *ALM, Level I*, 1961, and *Ecouter et Parler*, 1962
German, *ALM, Level I*, 1961, and *Verstehen und Sprechen*, 1963

DEFINITIONS OF TEACHING STRATEGIES

Traditional	Audiolingual + Grammar	Audiolingual
A. Use of native tongue in the classroom predominant. Target language not to be used for purposes of communicating instructions or information to students.	A. Use of target language in classroom. 1. By the student: for all responses. 2. By the teacher: for daily routine communication to pupils, instructions, cues and models.	A. Use of target language in classroom. 1. By the students: for all responses. 2. By the teacher: for daily routine communication.
B. Translation 1. Directly from native tongue to target language. 2. Reading by translation from target language to native tongue.	B. Native tongue to be used only for describing grammar and syntax.	B. Native tongue to be used only for describing grammar and syntax.
C. Vocabulary 1. Word for word equivalents. 2. Academic and literary lexicon stressed.	C. Sequence of learning 1. Hearing 2. Speaking 3. Reading 4. Writing	C. Sequance of learning 1. Hearing 2. Speaking 3. Reading 4. Writing
D. Grammar 1. Analysis before application. 2. Language organized into word lists, paradigms, principal parts, rules. 3. Analysis in depth of grammatical structures.	D. Grammar 1. Descriptive rather than prescriptive. 2. Incidental to functional skills being taught 3. Formally presented and discussed when appropriate.	D. Grammar 1. Descriptive rather than prescriptive. 2. Incidental to functional skills being taught.
E. Testing 1. Grades based on written tests.	E. Reading 1. Printed material always presented as a transcription of spoken forms. 2. As direct communication without the intermediary of translation from the target language to the native tongue.	E. Reading 1. Printed material always presented as a transcription of spoken forms. 2. As direct communication without the intermediary of translation from the target language to the native tongue.

DEFINITIONS OF TEACHING STRATEGIES
(continued)

2. Use of vocabulary and idiom quiz.
3. Frequent use of dictation test.
4. Use of tests requiring thorough knowledge of paradigms or lists.

F. Culture - the following cultural areas are emphasized:
1. Great historical and literary personalities.
2. Monuments.
3. Masterpieces of art, music and literature.

G. General orientation of traditional program is academic and intellectual.

F. Writing - learned first as a transcription of spoken forms.

G. Testing - written and oral tests given in order to test for listening comprehension and speaking proficiency as well as reading and writing skills.

H. Culture - "total culture" as reflected in language behavior is taught as opposed to refinement or prestige culture.

F. Writing - learned first as a transcription of spoken forms.

G. Testing - written and oral tests given in order to test for listening comprehension and speaking proficiency as well as reading and writing skills.

H. Culture - "total culture" as reflected in language behavior is taught as opposed to refinement culture.

LANGUAGE LABORATORY SYSTEMS

Tape Recorder (TR)

The simplest audio aid for the modern foreign language teacher is the classroom tape recorder. Its convenience and ease of operation as well as its low cost have made the tape recorder an integral part of the foreign language classroom even in schools equipped with more extensive facilities.

The inclusion of the classroom tape recorder alone as "laboratory strategy" reflected the insistence of many teachers that it was as effective as a more elaborate language laboratory. Statistically it represented the minimum baseline or "control" strategy.

Audio-Active Laboratory (AA)

This constitutes one type of "language laboratory." Each student position is equipped with a microphone, amplifier, and headset. Usually there is more than one tape recorder or other program source at the teacher console. Finally, the teacher console is wired for monitoring individual student performance.

The immediate and most cogent argument for this installation is the privacy and isolation afforded each student. Eliminating distracting noises is recommended if students are expected to discriminate new sounds that are distressingly similar to those of his own language and to other sounds in the foreign language.

It is also claimed that hearing his own voice following that of the tape master, with amplification of similar quality, allows for effective correction when there is disagreement. Multiple-program sources provide for small group instruction and facilitate flashbacks to previous lessons that require review.

Audio-Active-Record Laboratory (AR)

The addition of recording facilities at student positions provides the teacher with a significant tool in developing "functional" skills. Principally, the student records the master and his responses and then compares these during playback. Differing learning rates can be accommodated. This is a practical means for evaluating oral performances, and closer teacher supervision is possible than with less complete installations.

Competent language educators favoring the use of the record facility offer as support an argument based on the method by which language is learned. They claim that the learning of a foreign phoneme occurs as a result of conscious attention to the process of how it is produced. As a result, knowledge of the articulatory phonetics is a definite aid.

SUMMARY OF INSTRUCTIONAL CONTROLS

Experimental Design

1. "Real life" situation
2. Preferred design (NON-Equivalent Control Group, C&S No. 10)
3. Extensive pretesting
4. Sophisticated and conservative statistics: Analyses of Covariance & Tukey "A"
5. Random assignment
6. Two concurrent experiments (French and German)

Teacher

1. MLA Proficiency Tests
2. Experience parameters
3. Well qualified (average 10 years experience, 45 graduate hours, 62 per cent foreign travel and 40 per cent NDEA Institutes)
4. Large number (104)
5. Pre-experimental training
6. Quarterly evaluation meetings
7. Frequent irregular observation and rating for adherence to treatment

Method

1. Precisely defined
2. Laboratory treatment realistic
3. Detailed curriculum guides
4. Two distinct testing programs

Materials

1. Restricted to most widely used representative texts
2. No supplementary material permitted
3. Utilized commercial audio programs

Schools

1. Widely diverse and representative
    a. geography
    b. size
    c. socio-economic
2. Guaranteed cooperation
3. Only one treatment per school

Students

1. Regular enrollees
2. Repeaters and transfers excluded
3. Students with missing data dropped
4. Atypical (IQ, MLAT) classes deleted
5. Large numbers (2,171)

Tests

1. Program developed by specialists
2. Only standardized tests
3. Scorers trained at ETV

Reliability

1. Twenty-eight class, 700 student confirmatory replication

ORIGINAL HYPOTHESES

In order to arrive at conclusions related to the stated objectives the development of hypotheses, either expressed or unexpressed is necessary. Whatever the personal biases of the research personnel, the older "Traditional" approach was considered the control population and the newer "Functional Skills" audiolingual populations were the innovative experimental treatments. Logically, it is incumbent upon the newer, supposedly better, technique to demonstrate superiority in some form over the norm, the status quo. The challenger must bear the burden of proof.

Objective 1: To determine which of three teaching strategies is most effective in achieving each of the four foreign language objectives.

Hypothesis A: "Functional Skills" classes will achieve significantly higher than "Traditional" classes on the criterion measures of Listening Comprehension and Speaking (FS > TLM).

Hypothesis B: "Functional Skills" classes will equal "Traditional" classes in achievement on criterion measures of Reading and Writing (FS = TLM).

Hypothesis C: "Traditional" classes will score significantly higher than "Functional Skills" classes on 1939-41 criterion measures of Reading (translating), Vocabulary and Grammar (TLM > FS).

Objective 2: To determine which of three language laboratory systems is best suited, economically and instructionally, to the development of pronunciation and structural accuracy.

Hypothesis A: Classes using the language laboratory on a twice weekly schedule achieve significantly higher on criterion measures of Listening and Speaking (AA, AR > TR).

Hypothesis B: Classes in which students use the tape recorder achieve significantly higher on criterion measures of Listening and Speaking (AR > AA).

Objective 3: To determine the optimum combination of "strategy" and "system" in achieving the goals of the foreign language program.

Hypothesis: There exists some combination of instructional strategy and audio system in which students achieve significantly higher on criterion measures of Listening and Speaking.

Objective 4: To determine variables and combinations of variables which best predict student achievement on criterion measures.

Objective 5: To determine correlations among language skills.

Objective 6: To determine if "strategy" and "system" are related to student ability.

Hypothesis A: Students with above average ability will achieve significantly better in "Traditional" classes than peers in "Functional Skills" classes on criterion measures (High: TLM > FS).

Hypothesis B: Students with average and below average ability in "Functional Skills" classes will achieve significantly higher than peers in "Traditional" classes (Mid: FS > TLM; Low: FS > TLM).

Objective 7: To identify and compare student attitudes toward each of the teaching strategies and language laboratory systems.

A. Which teaching procedures in both the traditional and audiolingual approaches generate student interest?

B. Which factors motivate a student to study a foreign language?

C. To what degree do the audiolingual and traditional programs fulfill student expectations in language mastery?

Objective 8: To identify levels of foreign language mastery that are attainable in the secondary school language program.

A. Classes can reasonably progress through text materials at the rate implied or stated by the authors.

B. It is possible to develop local norms and levels of achievement expectation on standardized tests.

Objective 9: To determine the strengths and weaknesses of selected-commercial programs.

Hypothesis: Within each strategy, classes utilizing one set of materials will achieve significantly higher on criterion measures than students learning other materials (TLM: A > B > C; FS: A > B).

Objective 10: To identify teacher factors related to student achievement.

A. Teacher experience and education relate to their ability to impart foreign language skills to students, i.e., there exist relationships among teacher experience factors and student/class achievement on criterion measures.

B. Teacher proficiency ratings by self, by observer and by objective tests scoring relate to teacher ability to impart foreign language skills.

In summary, then, the most powerful demonstration of differences in instructional efficiency would be for the "Functional Skills" classes to clearly show their supposed ability to foster significantly greater student achievement in the audiolingual skills, listening and speaking, and at the same time to maintain equality of achievement in the graphic skills, reading and writing.

Criterion Test and Publication Date	Hypothesized
MLA Listening - 1963	FS > TLM
MLA Speaking - 1963	FS > TLM
MLA Reading - 1963	FS = TLM
MLA Writing - 1963	FS = TLM
Cooperative Reading (Trans.) - 1939-41	TLM > FS
Cooperative Vocabulary - 1939-41	TLM > FS
Cooperative Grammar - 1939-41	TLM > FS

This demonstration was to be based primarily on the *MLA Cooperative Classroom Tests*, "... designed to fill the need for evaluation in schools using the audiolingual approach" (Handbook, p. 12).

ANALYSES OF DATA

At the end of one year of instruction, twenty-eight discrete measures (page 5) and six attitude-opinion indices were complete for 2,171 students, largely in grades 9 and 10. An individual student for whom complete data was not obtained was eliminated from the experimental population. Several entire classes in which the teacher had been observed deviating from the assigned strategy were summarily dropped from the project.

Statistical analyses were completed at the Computer Science Centers of the University of Maryland and West Chester State College. The programs provided analyses of variance and covariance. Reanalyses were done with varying criteria, covariates, contrasts and ordering. Analyses of secondary objectives used an analysis of variance and Tukey "A" multiple range tests between ordered means. A significant contrast, the primary unit for statistical analyses was the intact class mean. The statistical analyses were run several times with varying contrasts and covariates. Obviously only a few of the more pertinent contrasts of the hundreds computed can be summarized in an abbreviated report.

FIRST YEAR CONCLUSIONS

Conclusions permitted by observation of first year data analysis were as follows:

Objective 1: Comparative effectiveness of the three teaching strategies.

A. At the end of one year of instruction in French and German, "Traditional" classes significantly exceeded "Functional Skills" and "Functional-Skills + Grammar" classes on the 1939 and 1941 *Cooperative French/German Test*.

B. "Traditional" classes did significantly better than both "Functional Skills" strategies on the final *MLA Cooperative Classroom Reading Test* as well as the other approaches on the *Listening Test*.

C. "Functional-Skills + Grammar" classes achieved significantly better than "Functional Skills" classes in two different measures of reading and a vocabulary test but only as well as FSM classes on other measures, including the "Grammar" section of the Cooperative French/German Tests.

D. In a ten per cent sample of the experimental population (French N = 205, German N = 138) the "Traditional" classes did significantly better than "Functional Skills" classes on the MLA Cooperative Classroom Writing Test.

E. In the same sample, "Traditional" classes did as well as "Functional Skills" classes on the MLA Cooperative Classroom Speaking Tests.

Objective 2: Comparative effectiveness of the three language laboratory systems.

A. The language laboratory systems employed had no measurable effect on achievement on tests of listening, reading, vocabulary or grammar after one year of French or German instruction.

B. In a random ten per cent sample of each class not employing a language laboratory but equipped with classroom tape recorders, "Traditional" classes did better than "Functional Skills" classes on the MLA Cooperative Classroom Speaking Test.

C. Laboratory type had no effect on Speaking Test scores.

Objective 3: Determine optimum strategy-system combination:

--None was detected in the experimental population.

Objective 4: To determine the best predictors of foreign language achievement.

A. There were significant relationships between intelligence, aptitude, attitude and student marks in other subjects and foreign language achievement.

B. The most significant combination of predictors were the Modern Language Aptitude Test, a foreign language Listening Test and the Language I.Q. for both languages in grades nine through eleven.

Objective 5: To determine the relationship among the four skills: listening, speaking, reading and writing.

--All skills were highly interrelated and also correlated significantly with listening discrimination and expressions of student attitude and interest.

Objective 6: To determine whether strategy and system relate to student ability.

A. Students achieved most in the "Traditional" strategy despite individual differences in ability.

B. Student achievement reflected ability rather than strategy.

C. Females had a significantly higher foreign language aptitude than males.

Objective 7: To identify and compare student attitude toward language learning.

A. Student expectations and orientation were still overwhelmingly traditional. Two-thirds of all students studied a foreign language for college entrance requirements. Ninety per cent of a random sample ($N = 300$) had an initial "Traditional" expectation for their foreign language study.

B. Students anticipated liking foreign language study and became less favorably inclined as the school year progressed. The rate of decline was the same during the first year regardless of the language studied or the strategy employed.

C. Females had a more favorable attitude throughout a year of foreign language instruction than males. Males studying German had a somewhat better attitude toward foreign language study than males studying French.

D. Initial attitude was not related to later achievement.

Objective 8: To determine levels of functional mastery.

A. Many students achieved meaningful scores on pre-instructional foreign language tests. This implies no "zero" starting point and makes suspect research based solely on final testing.

B. Authors and publishers of "Functional Skills" materials imply too high an expectation of progress through their programs.

Other Conclusions:

A. Females achieved better in foreign languages than males; on almost all measures, in all strategies, and in all grades included in the experimental population.

B. Project teachers were well prepared by current standards, averaging ten years of teaching experience and forty-five semester hours of graduate education.

C. Assessment of teacher proficiency by competent observers correlated highly with teacher scores on the MLA Proficiency Test for Teachers and Advanced Students. They did not correlate with teacher self-ratings.

D. Sex of the teacher had few significant effects on student achievement.

E. There was no significant relationship between scores of eighty-nine French and German teachers on all seven parts of the Teacher Proficiency Tests and the achievement scores, both gross and gain, of their classes in foreign language skills.

SECOND YEAR CONTINUATION AND REPLICATION

Fifty intact classes (1,090 students) maintained the experimental treatment through Level II French and German. Under additional funding a twenty-eight class (700 student) replication was completed of the first year using the same teaching strategies, texts and testing program.

Major objectives and conclusions of the experiment after two years of instruction and an adequate replication were as follows: (Tables containing summaries of appropriate statistical analysis are reproduced in Appendix A.)

1. To determine which teaching strategy among the traditional, audiolingual or modified audiolingual approaches best accomplishes the four objectives of the foreign language program in the secondary school-listening comprehension, speaking fluency, reading and writing.

   Conclusion: No significant differences existed among strategies on all skills except reading (TLM >) as measured on contemporary standardized tests after two years. "Traditional" classes had achieved significantly higher on 1939-41 measures of reading, grammar and writing by the end of Level I.

2. To determine which language laboratory system is most effective.

   Conclusion: The language laboratory of any type, used twice weekly, had no discernable effect on achievement.

3. To determine the best predictors of success in foreign language achievement.

   Conclusion: The best over-all predictors of success were prior academic success and the Modern Language Aptitude Test.

4. To identify student attitudes toward foreign language instruction.

> Conclusion: Student attitude toward foreign language study declined throughout instruction, independent of the teaching strategy employed.

5. To ascertain levels of language mastery.

> Conclusion: Published test "norms" and implied in-text layout progress were not realized by most of the experimental population.

6. To identify strengths and weaknesses of selected commercial texts.

> Conclusion: Within the functional skills strategies students utilizing Holt, Rinehart and Winston materials did significantly better than students using the Audiolingual Materials.

7. To identify teacher factors related to student achievement.

> Conclusion: Neither teacher experience in years and graduate education nor scores on the MLA Teacher Proficiency Tests were related to mean class achievement after either one or two years.

FIRST AND SECOND YEAR DISCUSSIONS AND IMPLICATIONS

Throughout the research, one goal was foremost in the minds of the staff: to evaluate curriculum trends in a school situation approaching the reality of secondary education in the United States. The research was never conceived as an original experiment but as the large scale replication of previous studies in a broader yet more relevant context.

One serious disadvantage affecting the interpretation of the research was the choice of the word "Traditional" rather than the semantically less loaded term "Cognitive Code-Learning" advanced by Carroll (1965). The two appear to be defined in very similar terms. Throughout the experiment each strategy was hopefully represented in its best possible manner. The "Traditional" strategy as employed in the research was far different from the typical foreign language classroom instruction of the 1920's and 1930's.

The research staff is aware of the tendency to assume that teachers deviated from their assigned teaching strategies as a rationalization of the lack of significant findings in favor of newer strategies and materials. A number of reasonable controls were exercised within the confines of the normal school routine.

The experiment was an improvement over previous in situ research in modern foreign languages in a large number of students representing two languages was involved in each treatment. Materials and tests were not specially written but were those most available and in widespread use. The statistical analyses were sophisticated and conservative. Data gathering was as extensive and meticulous as could be permitted. Reporting has attempted to be factual and objective despite the fact that the conclusions of the research are often in direct opposition to the professional training, biases and intuition of the reporters.

Perhaps the greatest implication inherent in the conclusions of Projects 5-0683 and 7-0133 is that the foreign language education profession has for the past decade or more been predicating teaching strategies, materials, and electro-mechanical devices on theoretical assumptions that may not be entirely valid. The implication for a reexamination of the theoretical basis for second language learning in the secondary school environment is evident in the research.

The false implication that foreign language teaching revert to "Traditional" classroom techniques of the 1930's can not be read into the research. "Traditional" teachers as defined in the research benefited from many more insights into human growth, personal interrelations and the learning progress than their predecessors of forty years ago.

Countless improvements have been made in the physical classroom, text format and arrangement, and curriculum development. The generation of students utilized in this research has always known television, traveled more widely and seen the world grow smaller. Neither the teacher, the school, nor the students are the same from year to year. Retrogression is not possible and cannot be regarded as an implication of the research. The recasting of theory, perhaps once adequate, into current society is implied.

The implication is clear that the "lock-step" language laboratory in the secondary school, no matter of what type, does not meet the expectations posited by earlier, more closely controlled research. The twice-weekly utilization employed in the research may not be optimal but reflects the typical school practice as determined by surveys conducted both before and after the research experiment.

The implications are obvious that student recording equipment may be too ambitious an investment for student drill and pattern practice and that the classroom tape recorder offers the advantage of the "lock-step" language laboratory at a fraction of the cost.

The lack of a demonstrable relationship between scores of the MLA Proficiency Tests for Teachers and class achievement implies that the most important phase of education is the process of teaching--not the teacher's background in subject matter. The research, in examining student attitude, superficial classroom methodology, and teacher proficiency may have failed to examine the real causes of variation in achievement. These may lie in the unexplored area of process--student motivation for second language learning and student-teacher interaction. The implication is that more precise examinations need to be made of the role of motivation and classroom interaction and second language learning.

"Audiolingual with Grammar" classes were felt by the project teachers themselves to be the probable "winner" on a poll taken at the end of the two year experimental phase. Such was not the case, rather the strategy in which grammar was presented first, then practiced seemed to be more effective. The implication is obvious for research on deductive, "grammar before," versus inductive, "grammar after," on large enough scale to be sufficiently generalizable.

FIRST AND SECOND YEAR RECOMMENDATIONS

In the light of the conclusions that must be drawn from the data, the reporters of the research make the following recommendations to the profession:

1. Since the results do not replicate other smaller scale studies ...

   A. There should be established a center for the continuing long-term study of modern foreign language instruction within the milieu of the "real school" environment, especially concerning itself with the transfer and replication of localized experiments into large scale, curriculum-changing research;

   B. A similar but more precise experiment should be undertaken involving the teaching of Spanish;

   C. That future research include more precise definitions of "traditional teacher" and "audiolingual teacher" based on detailed physical and verbal interaction analyses.

2. Experimental research design in foreign languages should always include extensive pretesting, including skills tests, to permit more meaningful statistical analyses.

3. Since teacher scores on the MLA Teacher Proficiency Tests had little to do with the class achievement...

   A. That research be undertaken to adequately determine the relationship between various levels of teacher proficiency and student achivement;

   B. That the MLA Teacher Proficiency Tests not be used as a major factor in the certification of teachers until their value has been more clearly established.

4. A foreign language Listening Comprehension test should be made an integral part of foreign language aptitude tests.

5. A sound policy of language laboratory administration and maintenance be immediately initiated by responsible school authorities.

6. Separate norms should be reported for males and females on standardized modern foreign language achievement tests.

7. Secondary schools should provide a classroom tape recorder for each foreign language teacher for daily use before equipping special electronic classrooms.

8. Language laboratories should be equipped with student recording facilities for testing purposes and individualized study rather than for frequent recording of regular drill sessions.

9. Detailed studies should be undertaken of the role of motivation in foreign language learning by secondary school students with emphasis on identifying possible points of departure for behaviorally oriented research.

10. The foreign language education profession should become more directly aware of the implications of research on the individual classroom at all levels.

In conclusion, the study of the relative effectiveness of various teaching strategies and language laboratory systems seems to point out that curriculum innovations in foreign language have been widespread but that this impact may have been more superficial than the profession had hoped. Certainly, more study is needed to advance knowledge of the second language learning process in the realistic setting of the public school.

SECTION II

THIRD AND FOURTH YEARS OF THE STUDY

INTRODUCTION

Project 7-0133 was funded in 1966 for the express purpose of continuing, replicating and expanding upon Project 5-0683, a large scale investigation of the relative effectiveness of three teaching strategies and three language laboratory systems. This project, now completed, has shown that newer methods and electromechanical aids are not as effective in actual school situations as had been supposed. The instructional phase of Project 7-0133, confirms these findings both by replication and extension.

Specifically the studies indicated: (1) the "Traditional" students exceeded or equaled "Functional Skills" students on all measures; (2) language laboratories employed twice weekly had no discernable effect on student achievement; (3) student attitudes toward foreign language learning are independent of the way in which he is taught; and (4) there is no relationship between teacher scores on all seven portions of the *MLA Foreign Language Proficiency Tests for Teachers and Advanced Students* and the achievement of their classes in foreign language skills.

Project 7-0133 was fortunate in that the Commonwealth of Pennsylvania became increasingly interested in participation in the direct support of the research. This increased support evidenced by the assumption of many of the costs originally assigned to federal funding by West Chester State College and the Department of Education, permitted the conservation of resources to extend the study longitudinally. This assessment of the typical secondary school foreign language program through advanced levels was a fundamental purpose of the study.

The extension of the modern foreign language sequence in the public schools has long been a major goal of the profession. Ample evidence of this can be seen in the movement toward foreign languages in the elementary schools and the six-year sequence (grades 7-12) endorsed by the Modern Language Association, the National Education Association, most state departments of education and a wide variety of other professional organizations. Pennsylvania has been a leader in this longitudinal expansion by mandating that "... a minimum of a four-year sequence of a modern foreign language shall be offered by each school system." Such a program was a prerequisite for selection of participating schools in Projects 50-683 and 7-0133.

The "Statement of the Problem" section of the original proposal for Project 7-0133 specifies the fundamental differences between introductory and advanced levels of modern foreign language instruction, each with distinctive purposes. These result from the differing philosophies regarding the objectives and strategies--and thus the classroom materials and techniques. For example, one publisher of a widely used "Functional Skills" text introduces his approach to the teaching of reading skills with the statement:

> ... Level One makes a careful distinction between two kinds of reading: (1) reading in the sense of pronouncing words and sentences aloud in response to the stimulus of a printed or written sentence and (2) reading for comprehension. Level Two is concerned with the development of the second type of reading. Its aim is to develop the ability... to read with understanding without translating. (Harcourt, Brace and World, Inc.)

Similarly, there are differences that distinguish the teaching of grammar, developing listening and speaking skills, and the instruction in writing at the two levels. One important purpose of the extension of Project 7-0133 was, then, to assess student achievement in mastery of those skills that are taught in Level II.

It has been obvious since the first publication of "Functional Skills" texts in the early 1960's that Level I and Level II do not coincide with the usual Year I and Year II in the school year. The typical class does not usually complete Level II until well into the third year of instruction (Smith and Baranyi, 1968). In order to assess Level II, then, it was imperative to continue to observe students through the third year of instruction.

Another, and perhaps more pervasive purpose of a continuing study was to provide longitudinal data on language learning in the setting of the typical secondary school. Education, in general, and mastery of a second language, in particular, are longitudinal processes; the appropriate manner in which they are to be studied should be longitudinal. Often initially dramatic results favoring one approach or another may prove premature when assessments are made over a long period of time.

No realistic study of the effects of the Pennsylvania four-year mandate has been undertaken, especially as the extended sequence pertains to individual student growth in the typical secondary school situation. Basic questions concerning the expected levels of proficiency, the early identification and motivations of continuing students, and student aspirations and expectations are unanswered. Of equal importance are the possible effects of early teaching strategies. Lastly, studies of the relationships between teacher factors and student achievement and motivation on an extended sequence basis have not been completed.

Carroll (1963) has pointed out that modern foreign language, with a nominal "zero" starting point, lend themselves well to educational research. While students do achieve meaningful scores on foreign languages tests prior to formal instruction (Smith and Berger, 1968), such exposure is certainly less than the pre-knowledge the student may have in many other areas of the curriculum.

Since the completion of extended sequences of foreign language study is generally agreed to be both necessary for mastery and a "good thing" in general, it was considered mandatory to utilize the great wealth of student data available from Projects 5-0683 and 7-0133 to examine the following specific objectives as they pertain to an extended foreign language sequence.

THIRD YEAR OBJECTIVES

1. To determine which of three foreign language teaching strategies is most effective in achieving the foreign language objectives, listening comprehension and reading skills.

2. To determine which of three language laboratory systems is best suited, economically and instructionally, to the development of audiolingual skills.

3. To determine which variable, or combination of variables - IQ, total grade point average, and appropriate prognostic test - best predicts student achievement in foreign languages in each of the four foreign language skills and in overall language mastery.

4. To identify and compare student attitudes toward each of the teaching strategies and language laboratory systems.

5. To identify teacher factors related to student achievement.

THIRD YEAR FOLLOW-UP

From its inception, the research study had stated as its objective the longitudinal observation of a number of secondary school foreign language students. During the latter part of the second year of the experiment, the decision was made to attempt to observe as many students as possible during their third year of foreign language study. It was evident that the high rate of attrition among both students and teachers precluded the continuation of strict experimental controls. Continued observation but not manipulation was possible.

Accordingly, over three hundred original project students in twenty-four classes were observed during French or German III. They continued following the basic course materials that they had utilized in Levels I and II. Since foreign languages suffer from a high attrition rate among students, it was also decided to investigate the reasons for the continuation or non-continuation of foreign language study.

The third year study, then, should be regarded as a "follow-up" evaluation of the experimental instruction rather than as a controlled study since strategy distinctions seem to become less obvious in advanced levels.

THIRD YEAR EVALUATION

Project students who remained available to the researchers were tested in September, 1967 and in May, 1968--the beginning and end of Level III instruction. The Fall, 1967, testing included:

1. *MLA Cooperative Classroom Listening Test*, Form L
2. *MLA Cooperative Classroom Reading Test*, Form L
3. *Student Opinion Scale*
4. The *Junior Index of Motivation*

At midyear the students were asked to complete a paper-pencil survey of their reasons for having continued to study a foreign language for a third year. At the end of the third year, students were again tested, this time with a new form of the tests:

1. *MLA Cooperative Classroom Listening Test*, Form M
2. *MLA Cooperative Classroom Reading Test*, Form M
3. *Student Opinion Scale*

The M-form of the achievement tests was used as a final measure due to the advanced level of the students and their familiarity with the L-form, given in preceding years. In general it proved to be still too difficult for most students even after three full years of foreign language study.

RESULTS OF THE COMPARISON OF TEACHING STRATEGIES, THIRD YEAR

Too few students (N=8) remained in the "Traditional" experimental treatment during French III to permit a valid comparison with students in "Functional Skills" classes. At the completion of German III, however, comparable groups of students still remained in each of the three teaching strategies. The distribution of these students among the teaching strategies is reported in Figure 2.

FIGURE 2

DISTRIBUTION OF GERMAN III STUDENTS

TLM	46

	TR	AA	AR
FSG	31	8	11
FSM	none	44	3

Initially, analyses of variance were computed to determine the significance of differences among the three strategies on the final _MLA Listening Test_ (MA) and the _MLA Reading Test_ (MA). The analyses, shown in Tables 1 and 2, indicate that in _Listening_ "Traditional" students achieved significantly higher than "Functional Skills Method" students ($p < .01$) who in turn outscored "Functional Skills + Grammar" students ($p < .01$). On the _MLA Reading Test_, "Traditional" students again achieved significantly higher scores than either of the "Functional Skills" groups ($p < .01$).

Since the preliminary analysis indicated significant differences did exist among the strategies (TLM>), analyses of covariance were computed. The following results are based on using individual student scores as the basis of statistical analysis:

TABLE I

ANALYSIS OF VARIANCE BY STRATEGY,
GERMAN III: FINAL MA LISTENING TEST

Strategy	N	Mean	S.D.
1. Traditional	56	17.68	5.35
2. Functional Skills + Grammar	63	14.08	5.13
3. Functional Skills	63	16.03	5.66

Source	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	386.81	193.41	6.667**
Within	179	5192.75	29.01	

SIGNIFICANCE OF DIFFERENCES BETWEEN ORDERED MEANS

Tukey "A" Multiple Range Test

Group	3.	1.
2.	1.95**	3.60**
3.		1.65**

** $p < .01$

TABLE 2

ANALYSIS OF VARIANCE BY STRATEGY,
GERMAN III: FINAL MA READING TEST

Strategy	N	Mean	S.D.
1. Traditional	56	16.39	6.10
2. Functional Skills + Grammar	63	12.71	3.36
3. Functional Skills	63	13.24	4.34

Source	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	461.22	230.61	10.548**
Within	179	3913.64	21.86	

SIGNIFICANCE OF DIFFERENCES BETWEEN ORDERED MEANS

Tukey "A" Multiple Range Test

Group	3.	1.
2.	.52	3.68**
3.		3.15**

** $p < .01$

ANALYSIS OF COVARIANCE, GERMAN III

Enough students remained in the experimental population through German III to permit meaningful statistical analysis of the influence of teaching strategy on achievement. Complete data for three full years was available for one hundred and forty-one German III students as follows:

Traditional:	4 classes, 46 students
Functional Skills + Grammar:	5 classes, 47 students
Functional Skills:	3 classes, 50 students

Since the particular computer program employed for the analysis of covariance required equal numbers of students per treatment, five randomly selected individuals were dropped from the FSG and FSM groups to equate them with the forty-six student traditional group.

Covariates chosen were the Language IQ measure of the California Test of Mental Maturity and the Modern Language Aptitude Test, both known to relate to foreign language achievement. Coefficients of correlation between covariates and criteria *MLA Cooperative Classroom Listening* and *Reading Tests*, Form MA, for the one hundred forty-three students are repeated below:

r : Language IQ, MA Listen = .298 p = .01
r : Language IQ, MA Read = .300 p = .01
r : MLAT, MA Listen = .342 p = .01
r : MLAT, MA Read = .107 p = .01

The analyses of variance for the covariates indicate that the three groups did not differ significantly in verbal intelligence. There was a highly significant difference ($p < .01$) in scores among the groups on the *Modern Language Aptitude Test*, with the "Functional Skills" students noticeably higher. This difference existed throughout the study.

The analyses of covariance are reported in subsequent Tables 3 through 6. On the *MLA Cooperative Classroom Listening Test*, MA, after three years of exposure to the dichotomous strategies, "Traditional" students achieved significantly better than their audiolingual counterparts ($p < .01$) although they were initially similar in verbal intelligence (Table 3) and despite a significant difference favoring "Functional Skills" students on the *Modern Language Aptitude Test*.

Readers will remember that the results of the analyses at the ends of Levels I and II indicated that German "Traditional" classes equaled "Functional Skills" on listening tests in both years but achieved significantly higher ($p < .05$) in reading at end of Level II. That "Traditional" should be significantly better on two analyses in both areas after Level III despite the specific emphasis of the "Functional Skills" approaches on audiolingual skills is somewhat unexpected.

TABLE 3

ANALYSIS OF COVARIANCE BY STRATEGY - GERMAN III

Traditional vs Functional Skills + Grammar vs Functional Skills

Criterion: Final MLA Cooperative Classroom Listening Test, MA
Covariate: Pre-Experimental Language IQ

		Means		
Strategies	N	Language IQ	MA Listen	Adjusted MA Listen
TLM	46	118.07	18.28	18.11
FSG	46	114.59	14.65	15.09
FSM	46	118.54	15.96	15.70
Grand	138	117.07	16.30	

Analysis of Variance of Covariate (Language IQ)

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	429.01	214.51	2.84
Within	135	10169.40	75.33	
Total	137	10598.41	77.36	

Analysis of Variance of Criterion

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	311.145	155.57	5.00**
Within	135	4203.67	31.14	
Total	137	4514.82	32.96	

Analysis of Covariance of Criterion

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	231.54	115.77	3.99*
Within	134	3890.50	29.03	
Total	136	4122.03	30.31	

* $p < .05$ TLM > FSG, FSM
** $p < .01$

TABLE 4

ANALYSIS OF COVARIANCE BY STRATEGY - GERMAN III

Traditional vs Functional Skills + Grammar vs Functional Skills

Criterion: Final MLA Cooperative Classroom Listening Test, MA
Covariate: Pre-Experimental Modern Language Aptitude Test

		Means		
Strategy	N	MLAT	MA Listening	Adjusted MA Listening
TLM	46	43.93	18.28	18.35
FSG	46	38.98	14.65	15.37
FSM	46	50.46	15.96	15.17
Grand	138	44.46	16.30	

Analysis of Variance of Covariate (MLAT)

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	3049.05	1524.53	7.42**
Within	135	27751.19	205.56	
Total	137	30800.24	224.82	

Analysis of Variance of Criterion

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	311.15	155.57	5.00**
Within	135	4203.67	31.14	
Total	137	4514.82	32.96	

Analysis of Covariance of Criterion

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	291.51	145.75	5.24**
Within	134	3725.86	27.81	
Total	136	4017.37	29.54	

** $p < .01$ TLM > FSG, FSM

TABLE 5

ANALYSIS OF COVARIANCE BY STRATEGY - GERMAN III

Traditional vs Functional Skills + Grammar vs Functional Skills

Criterion: Final MLA Cooperative Classroom Reading Test, MA
Covariate: Pre-Experimental Language IQ

		Means		
Strategy	N	Lang. IQ	MA Reading	Adjusted MA Reading
TLM	46	118.07	16.91	16.76
FSG	46	114.59	13.26	13.56
FSM	46	118.54	13.59	13.36
Grand	138	117.07	14.59	

Analysis of Variance of Covariate (Language IQ)

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	429.01	214.51	2.85
Within	135	10169.40	75.33	
Total	137	10598.41	77.36	

Analysis of Variance of Criterion

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	375.78	187.89	7.93**
Within	135	3199.67	23.70	
Total	137	3575.46	26.10	

Analysis of Covariance of Criterian

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	325.36	162.68	7.39**
Within	134	2950.90	22.02	
Total	;36	3276.26	24.09	

** $p < .01$ TLM > FSG, FSM

TABLE 6

ANALYSIS OF COVARIANCE BY STRATEGY - GERMAN III

Traditional vs Functional Skills + Grammar vs Functional Skills

Criterion: Final MLA Cooperative Classroom Reading Test, MA
Covariate: Pre-Experimental Modern Language Aptitude Test

		Means		
Strategy	N	MLAT	MA Reading	Adjusted MA Reading
TLM	46	43.93	16.91	16.93
FSG	46	38.98	13.26	13.45
FSM	46	50.46	13.59	13.38
Grand	138	44.46	14.59	

Analysis of Variance of Covariate (MLAT)

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	3049.05	1524.53	7.42**
Within	135	27751.19	205.56	
Total	137	30800.24	224.82	

Analysis of Variance of Criterion

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	375.78	187.89	7.93**
Within	135	3199.67	23.70	
Total	137	3575.46	26.10	

Analysis of Covariance of Criterion

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	378.79	189.39	8.01**
Within	134	3167.86	23.64	
Total	136	3546.65	26.08	

** $p < .01$ TLM > FSG, FSM

INFLUENCE OF PRIOR LANGUAGE LABORATORY EXPERIENCE

Analyses of variance were computed to determine if the type of language laboratory system that students utilized during Levels I and II had any discernible influence on achievement during Level III. Little meaningful information resulted due to the few students remaining in certain cells and the complete absence of students in some treatments. These results are reported in Tables 7 through 10.

Significant differences found between group means do not seem to follow a pattern. Since no significant differences between systems was found for Levels I and II with substantial numbers of students, significant differences among Level III groups are probably attributable to factors other than early laboratory treatment.

TABLE 7

ANALYSIS OF VARIANCE BY SYSTEM,
FRENCH III: Final MA LISTENING TEST

Strategy	N	Mean	S.D.
1. FSG-TR	9	18.67	4.15
2. FSG-AA	39	15.44	4.52
3. FSG-AR	31	14.06	6.29
4. FSM-TR	7	11.00	2.08
5. FSM-AR	9	15.67	7.48

Source	df	Sum Sqs.	Mean Sq.	F-ratio
Between	4	271.53	67.88	2.37
Within	94	2575.46	28.62	

TABLE 8

ANALYSIS OF VARIANCE BY SYSTEM, FRENCH III: FINAL MA READING TEST

Strategy	N	Mean	S.D.
1. FSG-TR	9	19.44	5.05
2. FSG-AA	39	14.97	5.42
3. FSG-AR	31	16.65	5.07
4. FSM-TR	7	12.71	2.50
5. FSM-AR	9	18.33	5.96

Source	df	Sum Sqs.	Mean Sq.	F-ratio
Between	4	284.42	71.11	2.65*
Within	90	2413.72	26.82	

SIGNIFICANCE OF DIFFERENCES BETWEEN ORDERED MEANS

Winer F Multiple Range Test[1]

Group	2.	3.	5.	1.
4.	2.26	3.93	5.62*	6.73*
2.		1.67	3.36	4.47*
3.			1.69	2.80
5.				1.11

* $p < .05$

---

[1] see Winer, op. cit., p. 100

TABLE 9

ANALYSIS OF VARIANCE BY SYSTEM, GERMAN III: FINAL MA LISTENING TEST

Strategy	N	Mean	S.D.
1. FSG-TR	32	15.50	5.44
2. FSG-AA	8	10.13	4.49
3. FSG-AR	23	13.48	4.14
4. FSM-AA	55	16.11	5.96
5. FSM-AR	8	15.50	3.17

Source	df	Sum Sqs.	Mean Sq.	F-ratio
Between	4	320.65	80.16	2.84*
Within	125	3419.96	28.26	

SIGNIFICANCE OF DIFFERENCES BETWEEN ORDERED MEANS

Winer F Multiple Range Test

Group	3.	1.	5.	4.
2.	3.35**	5.38*	5.38*	5.98*
3.		2.02	2.02	2.63
1.				.61
5.				.61

* p .05
** p .01

TABLE 10

ANALYSIS OF VARIANCE BY SYSTEM, GERMAN III: FINAL MA READING TEST

Strategy	N	Mean	S.D.
1. FSG-TR	32	13.47	3.46
2. FSG-AA	8	12.25	1.49
3. FSG-AR	23	11.83	3.52
4. FSM-AA	55	13.25	4.56
5. FSM-AR	8	13.13	2.53

Source	df	Sum Sqs.	Mean Sq.	F-ratio
Between	4	46.85	11.71	.775
Within	121	1828.08	15.11	

LEVEL III - "t" TESTS FOR INDEPENDENT SAMPLES FINAL MLA COOPERATIVE CLASSROOM LISTENING TEST, FORM MA

	Mean	S.D.	
French:			
FSG-AA (3 classes)	15.44	4.52	t = 1.02
FSG-AR (3 classes)	14.06	6.79	
German:			
FSG-TR (2 classes)	15.50	5.44	t = 1.57
FSG-AR (2 classes)	13.48	4.14	

PREDICTION OF ACHIEVEMENT, GERMAN III

Multiple regression equations were computed using pre-experimental (August-September, 1965) data on both teachers and students as predictors of student foreign language achievement at the end of German III. For this purpose student converted scores on the MLA Cooperative Classroom (Form MA) Listening and Reading Tests were added to give a composite measure of foreign language "achievement." This was done to provide a more meaningful and practical group of predictors than separate equations for each foreign language skill as was done for Projects 5-0683 and 7-0133.

Data on predictors is shown in Table 11 which illustrates the simple correlation coefficient between the nineteen predictors studied and foreign language "achievement" at the end of German III.

The teachers self-estimate of linguistic abilities correlated very significantly with student achievement as did teacher scores on the Listening and Reading Tests of the MLA Proficiency battery. Teacher scores on the Culture and Civilization and Professional Preparation achievement. Student verbal intelligence, aptitude and English achievement correlated significantly with subsequent foreign language "achievement."

The multiple regression equations themselves were computed separately by strategy and then for the entire student population. It is interesting to note in Table 12 that the single largest contributor to student achievement in each strategy group is different from that of the other groups and for the groups combined.

For the TLM and FSG groups the two best predictors include one teacher and one student measure. For FSM, student measures alone were the best predictors of later achievement. For German III students as a whole, teacher self-confidence in reading German combined with student verbal skills to maximize prediction of foreign language "achievement." Teacher scores on the Culture and Civilization Test enhances prediction when used in a negative manner. (Table 13)

TEACHER PROFICIENCY AND STUDENT ACHIEVEMENT

Projects 5-0683 and 7-0133 found that little significant relationship existed between pre-experimental measures of teacher proficiency and subsequent class achievement. Twelve German classes remained with the same teacher through Level III. These classes formed the basis of the analysis reported in Table 13. Transfer students were excluded from comparisons to permit the study of teacher-student relationships after three years.

TABLE 11

MEANS AND CORRELATIONS, PRE-EXPERIMENTAL VARIABLES AND SUBSEQUENT GERMAN III ACHIEVEMENT (N=102)

		Level III Achievement (MA Listen + MA Read)		
		Mean	S.D.	Correlation Coefficient
Teacher				
1.	Graduate hours	55.56	41.46	.393**
2.	For. Lang. Tchg. Experience	7.83	7.26	-.077
3.	1964 Salary	$6957.84	1639.19	.116
4.	Self-est., Speaking	2.29	1.09	.440**
5.	Self-est., Reading	2.73	.73	.644**
6.	Self-est., Writing	2.19	.64	.318**
7.	MLA Prof.: Listening	45.93	6.06	.230*
8.	Speaking	93.94	12.60	.170
9.	Reading	57.01	9.73	.199*
10.	Writing	61.54	12.24	.172
11.	Ap. Ling.	52.28	8.72	-.043
12.	Cult. & Civ.	57.18	5.48	-.251*
13.	Prof. Prep.	64.16	4.43	-.270**
Student				
15.	Lang. IQ	116.62	9.00	.448**
16.	Mod. Lang. Apt. Test	42.09	15.10	.245*
17.	Grade at start F.L. Study	10.88	3.34	.048
18.	Preceeding Eng. Grade	2.52	1.08	.329**
19.	Age, months	168.80	6.47	.074
Criterion:	MA Listen + MA Read Scores	311.17	16.11	

* p .05
** p .01

TABLE 12

MULTIPLE REGRESSION EQUATIONS
FINAL GERMAN III ACHIEVEMENT

Strategy: Traditional (N=37)

Coefficient	Variable	Beta	% Variance
- 3.88	Tchr. MLA Cult & Civ.	-.774	-58.10
+ 11.18	Tchr. MLA Prof. Prep.	1.404	109.22
+ .93	Stud. Lang. IQ	408	21.09
-265.38	Constant		

R = .85 F-test for significance = 61.68 (1,33)**

Coeff. Mult. Deter. = .722

Goodness of Fit, F=28.57 (3,33)**

F-test for addition of final variable (Tchr. MLA Cult & Civ.)
F=3.471 (1,33)

Strategy: Functional Skills and Grammar (N=34)

Coefficient	Variable	Beta	% Variance
.500	Tchr. MLA Reading	.342	17.30
+ .451	Student MLAT	.396	21.26
+261.998	Constant		

R-.621 F-test for significance = 6.98 (1,31)*

Coeff. Mult. Deter. = .386

Goodness of Fit = 9.73 (2,31)**

F-test for addition of final variable (Tchr. MLA Read)
F=4.88 (1,31)*

TABLE 12

(continued)

Strategy: Functional Skills (N=31)

Coefficient	Predictor	Beta	% Variance
.581	Student Lang. IQ	.325	13.53
.400	Student MLAT	.537	31.81
+216.897	Constant		

R = .673 F-test for significance = 11.89 (1,28)**

Coeff. Mult. Deter. = .477

Goodness of Fit = 11.610 (2,28)**

F-test for addition of final predictor (Lang. IQ) = 4.263 (1,28)*

Total Population: (n=102)

Coefficient	Predictor	Beta	% Variance
11.622	Tchr. Self-est. Read	.529	34.08
- .790	Tchr. MLA Cul & Civ	-.269	6.74
+ .556	Student Lang. IQ	.311	13.92
+ 2.632	Prior English grade	.176	5.81
+253.15	Constant		

R = .778 F-test for significance = 37.59 (1,97)**

Coeff. Mult. Deter. = .605

Goodness of Fit = 37.21 (4,97)**

F-test for addition of final predictor (Prior Eng. grade) = 5.786 (1,97)*

---

*p ∠.05
**p ∠.01

TABLE 13

TEACHER, STUDENT DATA AND CORRELATIONS BETWEEN TEACHER PROFICIENCY SCORES AND CLASS ACHIEVEMENT AFTER THREE YEARS

GERMAN III, 12 classes

Pre-Experimental Teacher Prof. Tests:	Mean	S.D.	Percentile[1]
1. Listening	43.58	8.15	62-65
2. Speaking	90.33	14.08	70
3. Reading	54.25	11.16	70-75
4. Writing	58.58	14.39	70
5. Applied Linguistics	52.42	8.71	70
6. Culture and Civilization	55.83	5.57	75-80
7. Professional Preparation	64.33	5.50	65

Post-Instructional MLA Cooperative Classroom Tests:

	As Individuals, N=181			As Intact classes, N=12		
	Mean	S.D.	Percentile	Mean	S.D.	Percentile
1. MA Listening	15.93	5.38	35	16.06	3.61	35
2. MA Reading	14.78	4.60	45	14.22	3.41	38

CORRELATION COEFFICIENTS[2]

	MLA Teacher Proficiency Tests:						
Class Achievement:	Listen	Speak	Read	Write	Ap. Ling.	Cult.	Pro. Prp.
1. MA Listen	.209	.139	.145	.142	.078	-.439	-.039
2. MA Read	.012	.197	.177	.268	.197	-.276	.161

---

[1] Pre-Institute percentile

[2] $r = .576$, $p = .05$

STUDENT OPINION CHANGES

An examination of student attitude toward foreign language study was made throughout the experiment. In Levels I and II student opinion of foreign language study declined steadily throughout the experiment but did not differ significantly among strategies.

Data showing the opinion shifts over a three year period by those students finishing German III are shown in Table 14 below:

TABLE 14

GERMAN STUDENT OPINION CHANGES, THREE YEAR PERIOD

	Traditional (N=45)		Func. Skill Gram. (N=49)		Func. Skill Met. (N=55)	
	Mean*	S.D.	Mean*	S.D.	Mean*	S.D.
1. Pre-Experimental	5.33	.74	5.53	.60	5.45	.65
2. After Level II	5.04	.72	5.30	.96	5.01	.98
3. After Level III	5.01	.98	4.65	1.01	4.82	1.14

Analysis of variance and Tukey "A" critical range tests indicate significant differences as follows:

By Administration:

TLM: Pre-exper. thru Level II, not sig.
Level II - Level III, $p < .05$

FSG: Pre-exper. thru Level II, not sig.
Level II - Level III, $p < .05$

FSM: Pre-exper. thru Level II, not sig.
Level II - Level III, $p < .01$

By Strategy:

Pre-Experimental: no sig. difference - TLM, FSG, FSM
After Level II : no sig. difference - TLM, FSG, FSM
After Level III : no sig. difference - TLM, FSG, FSM

---

*possible score ranged from a low of 1 to a high of 7

THIRD YEAR SUMMARY

In summary, a sufficient number of German students remained available to the project staff through Level III to support the conclusions drawn after Levels I and II: "Traditional" students equaled or significantly exceeded the achievement of "Functional Skills" students on the MLA Cooperative Classroom Listening and Reading Tests.

Using data from twelve German classes (N=102) who stayed with the same teacher for three full years, there still is a significant relationship between measures of teacher proficiency and the achievement of their classes.

Student opinion measures continued to show a downward decline consistent with trends from Level I and II but there continued to be no significant differences in student opinions among strategies.

FOURTH YEAR OF OBSERVATION

In order for a student to be observed through four full years of foreign language study, instruction must necessarily have begun in either grade 8 or 9. In addition, the student must have continued uninterrupted study within the same school building for the four years. Lastly, a project teacher must be willing to administer tests for both students and former project students in other Level IV classes.

Despite these restrictions a surprising number of project students were found completing Level IV classes. At midyear these students answered a questionnaire designed to provide insights into the reasons for continuing their study at advanced levels. Former project students and their classmates took the MLA Cooperative Classroom Listening and Reading Tests (MA). A few students took the Speaking and Writing Tests.

ANALYSIS OF COVARIANCE

Complete data extending over a full four-year period was obtained on ninety-two students, seventy-two German and twenty French. The German students were rather evenly distributed among three groups by strategy according to their early experimental treatment and subsequent materials bias:

"Traditional" (N=27)
"Functional Skills and Grammar" (N=21)
"Functional Skills Method" (N=24)

One student took one of the final tests, the MLA Cooperative Classroom Listening and Reading Tests, Form MA.

This sample permitted the computation of analysis of covariance using the pre-experimental Modern Language Aptitude Test as a covariate. Illustrated in Tables 15 and 16, these analyses indicate that significant differences existed pre-experimentally among the three groups (FSM>TLM>FSG) at the .01 level that were reflected in final achievement. MA Listening Test and Reading Test means vary in the same order, FSM>TLM>FSG. However, when adjusted for pre-experimental aptitude, the order becomes TLM>FSM>FSG in both languages but fails to reach a level of statistical significance. No significant differences existed among the three strategy groups on either criterion.

TABLE 15

ANALYSIS OF COVARIANCE BY STRATEGY - GERMAN IV

Traditional vs Functional Skills + Grammar vs Functional Skills

Criterion: Final MLA Cooperative Classroom Listening Test, MA
Covariate: Pre-Experimental Modern Language Aptitude Test

		Means		
Strategy	N	MLAT	MA Listening	Adjusted MA Listening
TLM	27	41.93	21.37	21.74
FSG	21	34.62	19.10	20.38
FSM	24	57.13	22.25	20.71
Grand	72	44.86	21.00	

Analysis of Variance: Covariate

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	6045.187	3022.594	14.99**
Within	69	13911.437	201.615	
Total	71	19956.625	281.079	

Analysis of Covariance

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	24.801	12.401	.256
Within	68	3290.021	48.383	
Total	70	3314.822	47.355	

** $p < .01$

TABLE 16

ANALYSIS OF COVARIANCE BY STRATEGY - GERMAN IV

Traditional vs Functional Skills + Grammar vs Functional Skills

Criterion: Final MLA Cooperative Classroom Reading Test
Covariate: Pre-Experimental Modern Language Aptitude Test

Means	MLAT	MA Reading	Adjusted MA Listening
TLM (N=27)	41.93	19.44	19.65
FSG (N=21)	34.62	16.76	17.60
FSM (N=23)	55.91	19.52	18.51
Grand (N=71)	44.30	18.68	

Analysis of Variance: Covariate

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	5222.125	2611.062	13.553*
Within	68	13100.687	192.657	
Total	70	18322.812	261.754	

Analysis of Covariance

Variation	df	Sum Sqs.	Mean Sq.	F-ratio
Between	2	49.729	24.865	.426
Within	67	3911.016	58.373	
Total	69	3960.745	57.902	

** $p < .01$

REGRESSION ANALYSIS

Both French IV and German IV student data were analyzed to ascertain the long range success of pre-experimental information as predictors of subsequent student achievement. Pre-instructional variables available were the student's age in months at the beginning of language study, his verbal IQ score from Part I of the *California Test of Mental Maturity* (short form), the *Modern Language Aptitude Test* (short form), and the pre-experimental administration of the *MLA Cooperative Classroom Listening Test*, Form LA.

The criterion for foreign language "achievement" was the sum of the student's converted scores on the MLA Cooperative Classroom Test, Form MA, Listening and Reading. These tests were administered in May, 1969, after four years of instruction.

The data (Table 17) indicated a marked difference in initial aptitude between students electing French and those electing German who continued the study through Level IV. German students averaged 44.75 on the MLAT and French students 52.35.

The analysis for French (N=20) indicated that the MLAT was the primary predictor of long-range success. However, the non-contribution of the language IQ factor and the small sample size indicate that the French analysis may be suspect.

In German, however, the results with a sample size of seventy-two is more acceptable. The German regression indicates that verbal intelligence was the highest contributor (13.58%) and that the Modern Language Aptitude Test was the second contributor (3.02%) to final achievement variance.

Examination of computed residuals indicates that both regression equations (coefficients and constants) are able to closely approximate real achievement despite the relatively low coefficients of multiple regression and multiple determination.

FOURTH YEAR SUMMARY

Level IV results support earlier findings that there is no advantage favoring "Functional Skills" classes in performance on tests designed to measure functional skills. IQ seems to be the best predictor of long-range student foreign language achievement within the secondary school setting.

FOURTH YEAR STUDENT VIEWS

In the final months of a four-year sequence of foreign language study, two hundred and fifty-two advanced French and German students each responded to a personal request from the project coordinator to complete a questionnaire of reasons for their decisions to continue foreign language study into advanced levels. The purpose of the questionnaire was to provide insight into student perceptions and to shed light on possible ways that concerned educators might encourage students to continue foreign language study.

The tabulation of student replies is shown in Table 18. All but twenty-eight of the two hundred and fifty-two students responding had received the majority of their foreign language instruction in an audiolingual "Functional Skills" approach.

TABLE 17

REGRESSION ANALYSES, LEVEL IV
CORRELATIONS

	French (N=20)						German (N=72)					
	Mean	S.D.	Correlation Coefficients				Mean	S.D.	Correlation Coefficients			
			2.	3.	4.	5.			2.	3.	4.	5.
Predictors:												
1. Age, months	166.55	11.33	-.07	.32	-.21	-.07	167.80	7.49	-.14	.10	.07	-.07
2. Lang. IQ	119.20	5.44		.50*	-.09	.10	117.41	8.07		.45**	.07	.30*
3. MLAT	52.20	11.58			-.07	.45*	44.75	16.86			1.0	-.10
4. Pre-Exper. List. Test	11.10	11.76				.20	12.06	7.52				-.06
Criterion:												
5. Final MA List. + Read.	33.25	1.59					31.58	2.96				

MULTIPLE REGRESSION
Criterion: Final MA Listen. + Read. (conv.) scores

French IV Predictors	Coefficient	Beta	% Variance	Coefficient	Beta	% Variance
1. Age, months	-.035	-.25	1.76	.012	.03	-.20
2. Lang. IQ	-.067	-.23	-2.35	.164	.45	13.58
3. MLAT	.090	.65	29.33	-.053	-.30	3.02
4. Pre-Exper. List. Test	.024	.18	3.54	-.025	-.06	.36
	+ Constant = 33.67			+ Constant = 29.96		

Multiple Correlation Coefficient = .57
Coefficient of Multiple Determination = .32

Multiple Correlation Coefficient = .41**
Coefficient of Multiple Determination = .17

* $p < .05$
** $p < .01$

TABLE 18

STUDENT VIEWS ON FOREIGN LANGUAGE STUDY

French and German (N=252)

SECOND LANGUAGE STUDY:		Number	% of Total
1. Have you studied a second foreign language in high school?	Yes	39	15.5
	No	213	84.5
2. Which second language have you studied?	French	9	3.6
	German	8	3.2
	Spanish	10	4.0
	Latin	10	4.0
	Other	2	.8
3. How many years of second language study?	One	13	5.2
	Two	16	6.4
	Three	9	3.6
	Four	1	.4
EXTENDED SEQUENCE:			
4. When was decision made to study foreign language for an extended sequence?	End of Level I	165	65.5
	End of Level II	56	22.2
	End of Level III	26	10.3
5. Did anyone encourage extended foreign language study?	Yes	124	49.2
	No	128	50.8
6. Who encouraged extended sequence?	Teacher	54	21.4
	Family	39	15.5
	Counselor	25	9.9
	Friend	3	1.2
	Other relative	3	1.2
	No response	128	50.8
7. Did anyone discourage extended foreign language study?	Yes	26	10.3
	No	225	89.3

TABLE 18

(Continued)

		Number	% of Total
8. Who discouraged extended sequence:	Teacher	2	.8
	Parent	6	2.4
	Friend	8	7.1
CURRICULUM:			
9. Why chose to study particular foreign language?	Family background	43	17.1
	Future studies, career	25	9.9
	To use: speak, read, travel	28	11.1
	Elem. School background	20	7.9
	Cultural background	54	21.4
	Advice of peers	17	6.8
	Avoid other languages	34	13.5
	Chance - no other choice	3	1.2
10. What future plans for foreign language?	College	88	34.9
	No use forseen	62	26.6
	Travel-study	35	13.9
	For. Lang. as a profession	11	4.4
	Other Profession	12	4.8
	Linguistic insights	12	4.8
	Reading	18	7.1
	Other	1	.4
11. Suggestions for improvement of foreign language experience:			
	More grammar, vocabulary, material	114	45.2
	More speaking	33	13.1
	Less memorization, oral repetition	23	9.1
	More cultural activities	13	5.2
	More homework	6	2.4
	Go faster	1	.4
	Go slower	5	2.0
	Better class control	6	2.4

The first three questions were designed to find out what percentage of Level IV students--presumably interested and talented--had studied a second foreign language. The responses (Item 1) indicated that fifteen per cent had studied two languages, most Item 3 for one or two years.

Student perception of their reasons for continuing into an extended sequence is reflected in Items 4 through 8. It is important to stress that student perceptions may be more important than actual fact since it is the perception that influences the individual decision to continue.

Item 4 indicates that two out of three Level IV students believed they decided at the end of Level I to continue foreign language study for several more years. However, only half the students felt that someone else had ever encouraged them to study foreign languages for several years (Item 5). Of this fifty per cent, only one-half again felt that a teacher had encouraged them to continue (Item 6).

These two items (5 and 6) reveal that of two hundred and fifty-two Level IV students, only one-quarter felt that a teacher had encouraged them to study the foreign language in depth. Few thought that someone had ever actively discouraged advanced study (Item 8).

Most advanced students thought they had made their original choice of French or German for purposes of expanding their cultural horizons (21%, Item 9). Seventeen per cent elected their language due to some sort of family background, either directly or romanticized ("My grandmother was German"). More students made a choice based on "avoidance motivation" (13.5%) than did so for either future studies (9.9%) or functional use (11%).

The largest proportion (34.9%) felt they would use their foreign language primarily for college entrance and requirements (Item 10). Fully one-quarter (26.6%) felt, after four years of study, that they could see no future use for the foreign language skills they had developed. About fourteen per cent foresaw travel or study abroad. Few (4.8% each) projected using their foreign language as a teacher or in other professional areas.

The final question asked of students was for their suggestions on how their foreign language experience could have been improved. Half felt that their courses should be more substantial, containing more structure, vocabulary and content material (45%). Thirteen per cent wished they had had more speaking emphasis. One student in ten reacted unfavorably to much memorization and oral repetition.

The students responses are indeed discouraging considering the number of pupils' completing the second year of language study; one in four felt encouraged by a teacher to continue into advanced levels; a third still had college requirements as their primary objective; one in four saw no real use for their language skills. On the positive side, fully half of the respondents felt that their courses should have been more substantial in content.

SUBSEQUENT IMPACT OF THE RESEARCH ON CURRICULUM AND LANGUAGE LABORATORY USAGE PATTERNS

In the Fall of 1968 former project teachers and school administrators were asked by the staff to answer questions concerning the impact of the research project on their school's curriculum and use of the language laboratory since the involvement of the school in the research. Sixty-seven of the 104 original teachers had remained in the school situation and were able to complete the questionnaire. Responses were also gathered from thirty-two school administrators who had served as the research project coordinator for their school district.

Illustrated are the responses from the teachers involved since these are judged more meaningful than those of the administrators.

FOLLOW-UP QUESTIONNAIRE

Please circle answers wherever possible

1. Did participation in the project result in changes in classroom methodology:

Yes	No	No Response
39 (58%)	23 (34%)	5 (7%)

2. Have any new foreign language text materials been adopted in the school in French or German since June, 1966?

Yes	No	No Response
40 (59%)	21 (31%)	6 (10%)

3. Have there been any permanent changes involving use of the language lab since September, 1966?

Yes	No	No Response
9 (13%)	53 (79%)	5 (7%)

I. Scheduling:
At present the language lab is used by each class at least:

a. once	weekly	19 (44%)	No laboratory or no response = 24
b. twice		19 (44%)	
c. three or more times		5 (7%)	

II. Maintenance and Repair of the Language Laboratory

a. The approximate age of the language lab since the date of installation is _____ years.

0-6 years	5
6-9 years	20
9+ years	12

b. In the school year 1967-68 the language lab was INOPERATIVE:

1)	never	of the time	20 (29%)
2)	1 - 10%		11 (16%)
3)	10-25%		19 (28%)
4)	more than 25%		9 (13%)
5)	more than 50%		6 (8%)

c. Does the school have a maintenance contract?

Yes	No	No Response
19 (44%)	24 (55%)	1 (2%)

d. A service man is called only in case of emergencies.

Yes	No	No Response
24 (35%)	21 (31%)	20 (29%)

e. The school makes its own repairs with staff assistance.

Yes	No	No Response
19 (28%)	25 (37%)	20 (29%)

f. Since September, 1966:

1)	Fewer mechanical problems have been found than previously.	6 (195%)
2)	There seems to have been no apparent difference.	25 (64%)
3)	The mechanical problems have increased noticeably	8 (20%)
	No response	26 (38%)

4. Have you heard educators (other than colleagues and those involved in the project) discuss the research?

a.	once	10 (14%)
b.	2 or 3 times	18 (26%)
c.	often, more than 3 times	3 (8%)
d.	no response	29 (43%)

5. What has been the reaction of the school regarding this project?

a.	favorable	17 (25%)
b.	unfavorable	5 (7%)
c.	no reaction	28 (41%)
d.	no response	12 (17%)

6. Did your participation in the project influence any of your colleagues in foreign language teaching?

Yes	No	No Response
26 (40%)	29 (45%)	12 (15%)

7. Did you personally benefit from participation in the project?

Yes	No
56 (89%)	11 (16%)

Thirty-nine, slightly less than two-thirds of the teachers responding, felt that participation in the research project had directly resulted in changes in methodology within their classrooms (Item I). The same number indicated the adoption of new materials for French and German instruction since the inception of the study (Item 2).

Fifty-three teachers or seventy-nine per cent indicated that there had been no change in language laboratory utilization since the conclusion of the first year of the research study (Item 3). Nineteen (44%) of those responding said their school's laboratory was used by foreign language classes once each week. An identical number (19-44%) were used twice weekly. Of the schools having language laboratories, then, *eighty-eight per cent* still used the language laboratory on a one or two times per weekly basis, within the minimal level investigated by the study--*despite the fact that research in which their school was directly involved in and reported to indicated that such utilization had no discernable effect on achievement*.

This seems to be a severe indictment of (1) the importance of the research as seen by participating schools; (2) of the lack of concern of curriculum planners for program evaluation and improvement; (3) of the inability of apprized persons to change the *status quo*; or, perhaps, (4) simply that participating educators never even read the reports and summaries sent to them of research in which they played an important role.

The majority of former project teachers reported that their language laboratories were more than six years old (32 of 37, 86.5%). Twelve of the thirty-seven (32.4%) were from six to nine years old. Some large percentage of older laboratories was expected since a laboratory installation had, after all, been one of the criteria for original inclusion in the experimental population in 1964. Such a high percentage (86.5) over six years old indicates that the life expectancy of laboratory installations may be higher than anticipated or that laboratories are not replaced after what would seem to be a substantial number of years of service.

Forty-six teachers responded to the item concerning the estimate of the amount of time the language laboratory was *inoperative* during the 1967-68 school year. Eleven (24%) of these felt the laboratory was inoperative never, that is, *always* operative. Nineteen (41%) saw it as inoperative from 1 to 10% of school time; nine (20%) estimated that their laboratory was inoperative from 10 to 25%. Six more than 25% of the time. One-third of the teachers responding, then, perceived the language laboratory in their schools as inoperative more than one day in ten.

IIc indicates that 55% of the schools did not have language laboratory maintenance contracts.

Very few (8-11%) of the teachers had heard the research project discussed three or more times by other professional educators in the year since the conclusion of the study. Twenty-five per cent had detected a favorable reaction to the research project, seven per cent

(5 teachers) perceived an unfavorable reaction. Forty teachers (58%) indicated their school had no reaction to the study (Item 6a). Forty per cent felt the experiment had influenced their colleagues in foreign language teaching. Slightly more (44.6%) felt that colleagues had been untouched by the study.

An overwhelming 72% (47) teachers believed that they personally had benefited from participation in the project. Three teachers chose not to respond and only six of the sixty-five responding felt that they had not gained by being involved as participants in the research.

Overall, teachers felt that the research study had been personally beneficial to them but the lack of change in the pattern of language laboratory usage indicates that the school itself had not benefited from one of the major conclusions of the study.

ADDITIONAL INFORMATION AND ANALYSES OF THE FIRST AND SECOND YEARS OF THE STUDY

SECTION III

Obviously, this part of the SUPPLEMENTARY REPORT can have little unity within itself as it will attempt to provide only additional bits and pieces of information omitted from the Final Reports of USOE Projects 5-0683 and 7-0133. This includes ERRATA, additional data analyses and reanalyses, and amplifications.

DEFINITIONS OF STRATEGIES

The specific criteria for each strategy were continually challenged by the former consultants to the project during the discussion conference and by other professionals at various times. Woodlen, who actually produced the criteria, was not a foreign language educator but a professor of educational research. Woodlen stated that he produced the definitions carefully from notes and tape recordings of the consultants' discussions.

The Educational Testing Service, in establishing norming populations for the MLA Cooperative Foreign Language Tests was content to classify participating classes more simply as "Traditional" or "Audiolingual" on the basis of a questionnaire completed by the school. The Handbook for tests reports:

> The criteria used for making these distinctions /traditional, audiolingual/ derived largely from information regarding the amount of time devoted to the foreign language, in the amount of time devoted to translation from one language to another, and the amount of time devoted to grammar discussions in English...

The authors of the Handbook point out that some difficulty arose in this type of classification and that there was an undefinable third group which was used as an independent equating sample.

While the term "Traditional" should have been avoided as semantically loaded, it may have been purposefully chosen for this very reason in light of the pre-experimental commitments of the Commonwealth. The "Traditional" approach as implemented seems the same as Carroll's (1965) "Cognitive Code-Learning" theory which he maintains is fundamentally different from the "Audiolingual" approach.

During the discussion conference, Berger pointed out the basic control imposed by the text materials. This is supported by Hanzeli (del Olmo, p. 19) when he states, "The package /A-LM/ as it exists has a certain built in emphasis."

In the same recent call for a reappraisal of foreign language methodology del Olmo (p. 27) writes:

> ...We should examine the list of characteristics of the audio-lingual approach that have been isolated by Rivers (1964), and Valdman (1966), and show how these characteristics fare in the pragmatic atmosphere of the classroom...

The Pennsylvania project attempted to do just this. At the inception of the study, definitions were regarded as adequate, precise and differentiating. At the conclusion, some did not perceive them to be so. One observer in the post study meeting stated his belief that the definitions would have been accepted as adequate and exemplary had the research but confirmed the pre-experimental biases of the profession.

TEACHER ABILITY AND PREPARATION

In the Final Reports of Projects 5-0683 and 7-0133 it was pointed out that teachers involved in the experimental instruction were those who were nominated by their administrators as "good" teachers and who indicated a willingness to abide by the restrictions imposed by the research design.

On several occasions persons anxious to explain away the findings of the study which fail to support newer approaches have rationalized that poorer teachers must have represented the "Functional Skills" approaches.

STATISTICAL COMPARISON

Statistical comparisons by analyses of variance on available information on participating teachers by strategy is summarized in Tables 19, 20, and 21. Table 18 shows that no significant differences existed among teachers in the three strategies (TLM, FSG, FSM) in either language in (a) graduate credit hours, (b) years of teaching experience, (c) years of language teaching experience, or (d) salary--usually a reflection of preparation, service, and longevity.

Tables 20 and 21 indicate that teachers in the three strategies had an equal estimate on their own ability to speak, read and write their foreign language, French or German.

A reasonable criticism of Projects 5-0683 and 7-0133's failure to find significant advantages for "Functional Skills" classes might be that teachers in these strategies were themselves deficient in "Audiolingual Skills" and thus could not foster this skill in their students. Despite the fact that the assumption that teacher proficiency influences student achievement may be itself a serious error. Table 20 shows that French teachers in "Functional Skills" classes scored higher than "Traditional" teachers on every one of the seven parts of the MLA Proficiency Tests for Teachers and Advanced Students. In five of the seven areas the differences are large enough to be significant: the critical Listening and Speaking measures ($p<.05$ FG >), Applied Linguistics and Civilization-Culture ($p<.05$ FS >), and in Professional Preparation ($p<.01$ FS >).

Differences on the Reading and Writing Tests are six to seven converted score points in favor of "Functional Skills" teachers but the resulting F-ratios fail to reach the required level of significance.

German "Functional Skills" teachers also scored higher than "Traditional" German teachers on all seven parts of the MLA Proficiency Tests although none of the differences was large enough to reach an acceptable level of significance. Converted score differences range from one to ten points between group means.

TABLE 19

COMPARISON OF TEACHER EXPERIENCE FACTORS

Teacher Factor	Group		Mean	F-ratio (Analysis Var.)
Graduate credits	French:	TLM	36.6	
		FSG	45.3	
		FSM	26.5	
	German:	TLM	38.5	.875 n.s.
		FSG	44.9	
		FSM	49.2	
Years Tchg. Exper.	French:	TLM	9.9	
		FSG	11.3	
		FSM	8.1	
	German:	TLM	11.8	.724 n.s.
		FSG	10.6	
		FSM	7.4	
Yrs. Tchg. For. Lang.	French:	TLM	6.4	
		FSG	8.0	.824 n.s.
		FSM	5.8	
	German:	TLM	6.2	
		FSG	8.4	1.508 n.s.
		FSM	4.2	
1964-65 Salary	French:	TLM	$6342	
		FSG	6289	
		FSM	5826	
	German:	TLM	6591	.345 n.s.
		FSG	5965	
		FSM	5778	

TABLE 20

COMPARISON OF TEACHER PROFICIENCY FACTORS BY STRATEGY, FRENCH

Factor[a]	TLM (N = 10) Mean	S.D.	FSG (N = 18) Mean	S.D.	FSM (N = 19) Mean	S.D.	F-ratio Analysis Var.
1. Speak	1.6	.84	1.6	1.1	1.7	.86	.132
2. Read	2.2	1.2	2.0	1.0	2.0	.98	.134
3. Write	2.0	.94	1.8	1.2	1.7	.91	.317
MLA Proficiency Tests							
1. Listen	33.4	6.1	39.0	8.0	41.4	6.7	4.60*
2. Speak	66.9	9.0	70.8	11.3	75.1	7.5	3.17*
3. Read	40.1	6.2	46.6	8.9	47.6	9.0	2.92
4. Write	40.4	8.7	46.9	9.4	47.4	8.1	2.51
5. Linguistics	43.6	7.6	49.7	8.2	51.9	6.5	4.58*
6. Cult. & Civ.	45.9	5.6	45.5	6.7	50.1	7.6	3.18*
7. Prof. Prep.	58.1	7.8	62.0	8.8	67.1	5.4	6.42**

*$p < .05$ at 2,55 df.
**$p < .01$ at 2,55 df.

A teacher self-rating--range of possible scores from 1 (Good) to 4 (Poor).

TABLE 21

COMPARISON OF TEACHER PROFICIENCY FACTORS
BY STRATEGY, GERMAN

	TLM (N = 6)		FSG (N = 18)		FSM (N = 19)		F-ratio
Factor a	Mean	S.D.	Mean	S.D.	Mean	S.D.	Analysis Var.
1. Speak	1.2	1.6	1.4	1.1	1.7	1.0	.574
2. Read	1.8	1.6	2.1	1.3	2.1	1.1	.113
3. Write	1.3	1.2	1.8	1.2	1.6	.9	.473
MLA Proficiency Tests							
1. Listen	37.7	6.6	42.6	8.1	40.1	8.4	.955
2. Speak	80.7	12.6	90.4	13.9	86.1	13.4	1.268
3. Read	47.8	10.5	52.1	10.2	48.3	10.0	.799
4. Write	49.8	15.0	59.2	12.8	52.6	11.0	1.900
5. Linguistics	47.8	11.1	54.1	6.4	52.6	7.2	1.584
6. Cult. & Civ.	49.0	5.4	53.6	9.8	50.4	7.7	.969
7. Prof. Prep.	62.3	7.4	63.1	6.5	62.4	6.9	.054

No F-ratios reported are significant

A teacher self-rating: 1 (Good) - 4 (Poor)

PRE-EXPERIMENTAL WORKSHOP

Prior to the beginning of the experimental instruction, all participating teachers were required to spend a week on the campus of West Chester State College for pre-experimental orientation and training. The adequacy of this period has been questioned by concerned professionals. To assist in understanding what this pre-experimental workshop entailed, a copy of the program is reproduced for informative purposes.

Sunday, August 22

4:00 Registration--Men's Dormitory
6:00 Dinner--Dining Hall
"Research and the Role of the Teacher"--
Dr. J. William Moore, Chairman, Department of Education,
Bucknell University

Monday, August 23

7:30 Breakfast
8:30 General Session--Choral Rehearsal Room, Swope Hall
"Research in Action"--
Dr. N. Sidney Archer, Director, Bureau of Research, D.P.I.
"Project 1330"--
Mr. Emanuel Berger, Research Associate, D.P.I.
10:30 Intermission
10:50 General Session
12:00 Luncheon
1:00 Seminar--Conditions No. 10 & 20[1] --Room 1, Swope Hall
All other conditions--Choral Room, Swope Hall
3:00 Intermission
3:20 Seminar--Assemble as at 1:00 P.M. above
5:30 Dinner
7:30 "Modern Languages: Teaching and Testing"--
Mrs. Mariette Reed, Professional Associate in Foreign Languages, Educational Testing Service

Tuesday, August 24

7:30 Breakfast
8:30 Teacher Assessment--Choral Room
10:30 Intermission
10:50 Teacher Assessment (continued)
12:00 Luncheon
1:00 Seminar--Conditions 10 & 20, Room 1
All other conditions, Choral Room
3:00 Intermission
3:20 Language Seminar--Condition 10, Room 1
Condition 20, Room 3
Condition 11-16 inclusive, Room 5
Condition 21-26 inclusive, Room 8
5:30 Dinner
7:30 "Foreign Language Testing"--
Eugene Hogenauer, Westtown School, MLA Test Development Committee

Wednesday, August 25

7:30 Breakfast
8:30 Teacher Assessment--Choral Room
10:30 Intermission
10:50 Language Seminar--Conditions 10 & 20, Room 1
Conditions 11-16 inclusive, Room 5
Conditions 21-16 inclusive, Room 8

---

[1] Condition Codes Key: 1st digit: 1 = French, 2 = German
2nd digit: 0 = TLM, 1 = FSG-TR, 2 = FSM-TR,
387 = FSG-AA, 486 = FSM-AA,
5 = FSG-AR, 6 = FSM-AR.

12:00 Luncheon
1:00 Laboratory II--Condition 10, Room 1, Swope Hall
    Condition 20, Room 3, Swope Hall
    Conditions 11-16 inc., Room 120, Recitation Hall
    Conditions 21-26 inc., Room 419, Henderson High School
2:30 Intermission
2:45 Laboratory III--Assemble as in Laboratory II
5:30 Dinner
7:30 Tour of Longwood Gardens, duPont Estate, Kennett Square
    Fountain Display at 9:00 P.M.

Thursday, August 26

7:30 Breakfast
8:30 Laboratory IV--Assemble as in Laboratory II
10:00 Intermission
10:15 Laboratory V--Assemble as in Laboratory II
12:00 Luncheon
1:00 Methods Seminar--Condition 10, Room 1, Swope Hall
    Condition 20, Room 3, Swope Hall
    Conditions 11, 13 & 15, Room 5, Swope Hall
    Conditions 12, 14 & 16, Room 6, Swope Hall
    Conditions 21, 23 & 25, Room 7, Swope Hall
    Conditions 22, 24 & 26, Room 8, Swope Hall
3:00 Intermission
3:20 Methods Seminar--Assemble as at 1:00 P.M. above
5:30 Dinner
7:30 "Foreign Language in the United States--Past, Present, and
    Future"--Dr. Kenneth W. Mildenberger, Director of Programs, MLA

Friday, August 27

7:30 Breakfast
8:30 Testing Policy and Procedure--Choral Room
9:45 Field Consultants Conference--Group A, Room 1
    Group B, Room 5
    Group C, Room 8
    Group D, Choral Room
10:30 Intermission
10:50 General Session--Choral Room
12:00 Luncheon

SUBSEQUENT PROFESSIONAL STATUS OF TEACHERS

It cannot be said that on the basis of available objective information that "Functional Skills" teachers were "inferior." If anything, the reverse would be true. On a subjective plane, a check on the professional status of former "Functional Skills" participants as held by their colleagues will reveal a high proportion of "very good" teachers.

At the time of writing former representatives of the "Functional Skills" strategy enjoy great professional esteem: several are employed by West Chester State and other colleges as Master Teachers in student-teaching situations; one is a leader in the Philadelphia Chapter of the American Association of Teachers of German, another similarly in the Western Pennsylvania AATF: one is completing an advanced leadership NDEA Institute abroad; another teaches a college methodology course; lastly, one now is a state supervisor of foreign languages.

To date, all teachers maintain that they did an honest professional job in following their assigned instructional approach. None of the teachers know neither if, nor why, his class may have been deleted from the experimental population.

INFLUENCE OF TEACHER NDEA INSTITUTE TRAINING ON CLASS ACHIEVEMENT

Among the data available to the project was the information that forty percent of the teachers involved had attended National Defense Education Act Institutes prior to the commencement of the experimental instruction. This proportion, twice the state average, indicates at the least an increased awareness on the part of the teacher toward recent curriculum changes.

Analyses of variance were computed to determine if such training seemed to differentiate the achievement of the classes of these teachers from those of teachers who had not benefited from such an experience. Teachers represented all experimental cells, permitting the comparison across strategies and systems and randomizing student variables.

Table 22 indicates no significant differences in achievement on the *MLA Cooperative Classroom Listening* and *Speaking Tests* between the classes with NDEA-trained teachers and those classes without NDEA-trained teachers. Starr (see Section IV) specifically warns that an assumption that NDEA-Institute training automatically means better teaching is fallacious. Institutes varied widely in level, in emphasis, and in effectiveness. Often poorly prepared teachers participated while better teachers did not.

The results of the analysis of variance support Starr's contention that the NDEA-Institute background does not *per se* indicate greater skill on imparting foreign languages to their students.

MLA COOPERATIVE CLASSROOM TESTS

In recent critiques and discussions concerning the research project, the use of the *MLA Cooperative Classroom Tests* as criteria for student achievement has been questioned. This is the thesis of Valette and Lado (Section IV). Lado states that he believes that the *MLA Cooperative Classroom Tests* were not precise enough to determine significant differences favoring the "Functional Skills" strategy.

TABLE 22

INFLUENCE OF TEACHER N.D.E.A. INSTITUTE TRAINING ON LATER CLASS ACHIEVEMENT

(French I)

	Final MLA Cooperative Classroom Listening Test, LA				
	N	Mean	S.D.	Percentile	t
1. Classes, Tchrs. w/NDEA training	20	14.79	2.95	51	.85
2. Classes, Tchrs. w/o NDEA training	40	14.44	3.06	51	
	Final MLA Cooperative Classroom Speaking Test, LA				
	N	Mean	S.D.	Percentile	t
1. Classed, Tchrs. w/NDEA training	20	25.63	9.49	31	1.13
2. Classes, Tchrs. w/o NDEA training	40	28.62	9.10	45	

Obviously, the tests in question are not perfect. It is equally obvious that the state-of-the-art in test construction and analysis has improved in the period 1964-1969. Critics must remember that in 1964 the tests were new, hailed as exemplary and thought by many leading professionals to be the long awaited tests that would indeed support new approaches to foreign language teaching.

The *Handbook* for the *MLA Cooperative Foreign Language Tests* points out that, "the tests are designed to measure the language skills in a functional context" and "...have been designed to fill the need for evaluation in schools using the audiolingual approach..."

When the audiolingual approach was attaining its initial popularity, it was obvious that students who learned from this type of instruction would not be able to score as well on extant standardized tests written to measure primarily reading skills and grammatical knowledge. From an empiric point of view the new approach was not defensible and proponents of functional approaches had to wait the development of tests with a new orientation.

In 1963 the profession produced the *MLA Cooperative Foreign Language Tests* developed under the direction of Nelson Brooks. These were hailed as "New Tests for a New Key" (Bryan) and accepted with confidence by concerned professionals as evidence in the 1964 Northeast Conference Report on *Ideals and Practices*:

> ...Successful teaching stands helplessly before inquiring administrator and irresponsible critic alike, unable to offer any reasonable proof that it is doing what it says it is. Fortunately for our profession, the instrument which makes evaluation possible is now at hand--the *Modern Language Cooperative Classroom Tests*...
>
> The committee urges the widest possible use of this testing program as an effective answer to a frustrating problem (p. 35).

As recently as 1966, Brooks addressed educational leaders through the *Phi Delta Kappan*:

> Up to the present, what is called the new approach is largely an act of faith. Research to prove the validity of its basic principles is scanty...mainly because the scientific measurement... is extremely difficult, and because the needed instruments have, up to now, not been available (March, 1966, p. 359).

Therefore, in selecting the *MLA Cooperative Classroom Tests* as the major evaluative instrument for the Pennsylvania Project during the 1963-64 planning period, the research designers assumed that the tests were the best available. Other researchers have since worked under the same general assumption for the literature reports many studies which have used the MLA tests as final measures.

STUDENT GRADE PLACEMENT

The placement of project students by grade is not entirely clear in the original reports, especially for the replication population. Grade placement for finishing students, those for who complete data were obtained and thus included in the statistical analyses, were as follows:

	Original		Replicators	
	French	German	French	German
8th	50	--	62	24
9th	680	524	145	111
10th	270	176	132	66
11th	232	186	54	68
12th	15	--		

ADDITIONAL ANALYSIS OF LEVELS I AND II

The primary statistical analysis of USOE Projects 5-0683 and 7-0133 had employed a Multivariate Analysis of Covariance program (MANOVA) with as many as six covariates, both pre-experimental and semester measures. Midyear foreign language measures were used as one of several simultaneously applied covariates (5-0683, Sections III-1 and Appendix D; 7-0133, Section III-2). This was done to provide every opportunity for fairness to each strategy in view of the extended pre-reading period of the "Functional Skills" approaches. Such analyses were intended to reduce the "shock" effect of tests on students from treatments that kept printed material from students for a period of weeks or months and reduced the advantage longer contact with reading may have had on "Traditional" classes.

The project staff has been repeatedly questioned about the wisdom of such analyses since it knowingly reduced early treatment effects and, in essence, reduced the comparison of Level I to one of from January to May, 1966, and of Level II from January, 1966, to May, 1967. Authors of the reports are often asked if analyses without midyear measures as covariates would have produced different results. The answer is affirmative.

Analyses of covariance were computed for the full two-year period using only pre-experimental measures as covariates. The most complete such analysis of covariance is summarized in subsequent tables. The covariate is the pre-experimental Modern Language Aptitude Test (Short Form) since it partially accounts for possible sex factors that might overshadow other measures such as verbal intelligence scores. Utilization of the MLAT as a single covariate also permitted the inclusion of a class originally dropped from the MANOVA program due to missing pre-experimental aptitude test scores.

The unit from the analyses of covariance was the class mean. Preliminary analyses of variance indicated that "Traditional" classes in French (Table 23) scored significantly less on the MLAT ($p. < .01$) than "Functional Skills" classes. This was not true among the strategies in German (Table 24).

German I reanalyses show results similar to French with more significant differences favoring the "Traditional" classes over the "Functional Skills" classes. Most surprising is the significantly higher achievement of "Traditional" classes on the MLA Speaking Test (French $p. < .05$).

Students for the 10% speaking sample were randomly selected and tested individually in extra-class situations by the project staff using identical tape recorders to insure uniformity of recording. Scorers were trained at the Educational Testing Service (USOE 5-0683, p. 39).

TABLE 23

ANALYSES OF COVARIANCE BY STRATEGY

French I (58 classes)

I. Analysis of Variance for Pre-measure: *Mod. Lang. Aptitude Test*

Variation	df	Sum Sqs.	Mean Sq.	F
Between	2	64667	32333.50	5.41**
Within	55	328785	5977.91	
Total	57	39352	6902.66	

Group Means:	MLAT
TLM (10 classes)	39.55
FSG (23 classes)	45.84
FSM (25 classes)	49.04

II. Analysis of Covariance: Criterion, *MLA Classroom List. Test* (LA)

Variation	df	Sum Sqs.	Mean Sq.	F
Between	2	222.09	111.04	.165
Within	54	36368.23	673.49	
Total	56	36590.31	653.40	

Group Means: *MLA Listening Test*

	Original	Adjusted
TLM	13.72	15.07
FSG	14.45	14.52
FSM	15.11	14.52

III. Analysis of Covariance: Criterion, *MLA Classroom Spk. Test*[1]

Variation	df	Sum Sqs.	Mean Sq.	F
Between	2	6967.29	3483.65	4.93*
Within	54	38157.89	706.63	
Total	56	45125.19	805.81	

Group Means: *MLA Speaking Test*

	Original	Adjusted
TLM	32.36	35.04
FSG	24.57	24.70
FSM	29.20	28.01

TABLE 23

(Cont'd)

IV. Analysis of Covariance: Criterion, MLA Classroom Reading Test

Variation	df	Sum Sqs.	Mean Sq.	F
Between	2	6647.64	3323.52	3.91*
Within	54	45907.84	850.15	
Total	56	52554.88	938.48	

Group Means: MLA Reading Test

	Original	Adjusted
TLM	16.60	17.90
FSG	15.37	15.42
GSM	15.14	14.56

V. Analysis of Covariance: Criterion, MLA Classroom Writing Test

Variation	df	Sum Sqs.	Mean Sq.	F
Between	2	24499.18	12249.59	10.86*
Within	54	60899.56	1127.77	
Total	56	85398.74	1524.98	

Group Means: MLA Speaking Test

	Original	Adjusted
TLM	32.69	36.14
FSG	18.82	18.98
FSM	18.06	16.54

---

[1] 10% random sample of each class

* $p < .05$
** $p < .01$

The reanalysis of French I gives results somewhat different than the data reported in USOE 5-0683, with more significant differences in favor of the "Traditional" approach. The reanalysis for German I is as follows:

TABLE 24

ANALYSIS OF COVARIANCE BY STRATEGY

German I (43 classes)

I. Analysis of variance for Pre-measure: Mod. Lang. Aptitude Test

Variation	df	Sum Sqs.	Mean Sq.	F
Between	2	53.00	26.50	.003
Within	40	335230.00	8380.75	
Total	42	335283.00	7982.93	

Group Means	MLAT
TLM (6 classes)	46.17
FSG (18 classes)	46.50
FSM (19 classes)	46.40

II. Analysis of Covariance: Criterion, MLA Classroom List. Test

Variation	df	Sum Sqs.	Mean Sq.	F
Between	2	1634.33	817.26	1.23
Within	39	25842.41	662.63	
Total	41	27476.73	670.16	

Group Means: MLA Listening Test

	Original	Adjusted
TLM	16.62	16.63
FSG	14.81	14.80
FSM	15.61	15.61

III. Analysis of Covariance: Criterion, MLA Classroom Speaking Test

Variation	df	Sum Sqs.	Mean Sq.	F
Between	2	2889.06	14449.03	1.25
Within	39	450851.75	11560.30	
Total	41	479749.81	11701.22	

Group Means: MLA Speaking Test

	Original	Adjusted
TLM	29.67	29.67
FSG	22.08	22.07
FSM	22.31	22.31

TABLE 24

(Cont'd)

IV. Analysis of Covariance: Criterion, MLA Classroom Reading Test

Variation	df	Sum Sqs.	Mean Sq.	F
Between	2	5583.65	2791.82	3.53*
Within	39	30875.10	791.41	
Total	41	36448.75	888.99	

Group Means: MLA Reading Test

	Original	Adjusted
TLM	17.22	17.22
FSG	13.71	13.71
FSM	14.77	14.77

V. Analysis of Covariance: Criterion, MLA Classroom Writing Test[1]

Variation	df	Sum Sqs.	Mean Sq.	F
Between	2	115521.00	57760.50	2.45
Within	39	921497.56	23628.14	
Total	41	1037918.56	25293.13	

Group Means: MLA Writing Test

	Original	Adjusted
TLM	39.97	40.06
FSG	24.24	24.21
FSM	26.61	26.62

---

[1] 10% random sample of each class

* $p. < .05$
** $p. < .01$

TEST SCORER RELIABILITY

Since the Directions for Administration and Scoring booklet for the MLA Cooperative Classroom Tests published by the Educational Testing Service makes the specific comment that the Speaking Test suffers from scorer reliability, it was deemed wise to check inter-scorer reliability.

In order to test uniformity of scoring on the important MLA Cooperative Classroom Speaking Test, even after receiving training and being checked by Educational Testing Service personnel, a statistical comparison was made of randomly selected tests scored independently by the two field consultants for each language. The comparison (French N = 64, German N = 18) demonstrates a significant correlation between the individual scorers. One French scorer marked higher than the other but this pattern was consistent as reflected by the highly significant correlation coefficient.

It should be noted that each scorer marked one-half the classes representing each experimental cell and that classes assigned to one scorer did not dominate an experimental treatment. The results of this analysis follow in Table 25.

TABLE 25

SCORER RELIABILITY, SPEAKING TEST

MLA Cooperative Classroom Speaking Test

French, Form LA: Independent scoring of 64 randomly selected tests:			
	Mean	S.D.	Correlation
Scorer A	29.16	10.67	.62**
Scorer B	23.95	10.30	
German, Form LA: Independent scoring of 18 randomly selected tests:			
	Mean	S.D.	Correlation
Scorer A	24.56	9.28	.47*
Scorer B	24.06	10.16	

* $p < .05$
** $p < .01$

ANALYSIS OF COVARIANCE BY TEACHING STRATEGY

Level II analyses of covariance were completed for twenty-four French and twenty-six German classes according to the teaching strategy. The analyses of variance for the covariate, the Modern Language Aptitude Test, indicate no significant differences among treatments for either language. Significant differences on post-measures in general support the analyses of Level I but with less significance appearing, particularly among German II classes. The results of these analyses are summarized in Table 26 (French) and Table 27 (German).

TABLE 26

ANAYLSIS OF COVARIANCE BY TEACHING STRATEGY

for French II Classes (N = 24)

I. Analysis of Variance for Pre-measure: Mod. Lang. Aptitude Test

Source	df	Sum Sqs.	Mean Sq.	F
Between	2	104.99	52.50	.79
Within	21	1399.98	66.67	
Total	23	1504.97	65.43	

Group Means:	MLAT
TLM (4 classes)	42.33
FSG (14 classes)	46.70
FSM (6 classes)	48.91

II. Analysis of Covariance: Criterion - MLA Classroom List. Test (LB)

Source	df	Sum Sqs.	Mean Sq.	F
Between	2	6.292	3.146	.233
Within	20	281.888	14.094	
Total	22	288.180	13.099	

Group Means: MLA Listening Test

	Original	Adjusted
TLM	21.01	22.58
FSG	21.30	21.23
FSM	21.93	21.04

TABLE 26

(Cont'd)

III. Analysis of Covariance: Criterion - MLA Classroom Speak. Test (LB)[1]

Source	df	Sum Sqs.	Mean Sq.	F
Between	2	487.619	243.810	2.808
Within	20	1736.718	86.836	
Total	22	2224.337	101.106	

Group Means: MLA Speaking Test

	Original	Adjusted
TLM	36.12	39.08
FSG	27.37	27.24
FSM	35.19	33.51

IV. Analysis of Covariance: Criterion - MLA Classroom Read. Test (LB)

Source	df	Sum Sqs.	Mean Sq.	F
Between	2	127.566	63.783	5.048*
Within	20	252.731	12.637	
Total	22	380.298	17.286	

Group Means: MLA Reading Test

	Original	Adjusted
TLM	25.53	26.90
FSG	20.74	20.68
FSM	21.00	20.22

V. Analysis of Covariance: Criterion - MLA Classroom Writ. Test (LB)

Source	df	Sum Sqs.	Mean Sq.	F
Between	2	2230.111	1115.056	5.635*
Within	20	3957.326	197.866	
Total	22	6187.437	281.247	

Group Means: MLA Writing Test

	Original	Adjusted
TLM	54.04	59.52
FSG	33.31	33.09
FSM	47.05	43.93

---

[1] 10% random sample of each class

* $p. < .05$

The German II classes also had no significant differences among strategies on the Modern Language Aptitude Test but several significant analyses favoring the "Traditional" strategy, although not to the same degree as in Level I. These differences can be found in Table 27.

TABLE 27

ANALYSIS OF COVARIANCE BY TEACHING STRATEGY

for German II classes (N = 26)

I. Analysis of Variance for Pre-measure: Mod. Lang. Aptitude Test

Source	df	Sum Sqs.	Mean Sq.	F
Between	2	86.016	43.008	0.382
Within	23	2586.363	112.451	
Total	25	2672.379	106.895	

Group Means:	MLAT
TLM (6 classes)	46.21
FSG (9 classes)	43.80
FSM (11 classes)	47.97

II. Analysis of Covariance: Criterion - MLA Cooperative List. Test (LB)

Source	df	Sum Sqs.	Mean Sq.	F
Between	2	0.440	0.220	0.011
Within	22	440.213	20.010	
Total	24	440.653	18.361	

Group Means: MLA Listening Test

	Original	Adjusted
TLM	19.59	19.58
FSG	18.93	19.23
FSM	19.67	19.43

III. Analysis of Covariance: Criterion - MLA Classroom Speak. Test (LB)[1]

Source	df	Sum Sqs.	Mean Sq.	F
Between	2	302.713	151.356	1.665
Within	22	2000.291	90.922	
Total	24	2303.004	95.958	

TABLE 27

(Cont'd)

Group Means: MLA Speaking Test

	Original	Adjusted
TLM	39.37	39.36
FSG	34.68	34.99
FSM	30.88	30.63

IV. Analysis of Covariance: Criterion - MLA Classroom Read. Test (LB)

Source	df	Sum Sqs.	Mean Sq.	F
Between	2	130.170	65.085	3.248
Within	22	440.785	20.036	
Total	24	570.956	23.790	

Group Means: MLA Reading Test

	Original	Adjusted
TLM	21.77	21.76
FSG	16.17	16.44
FSM	16.67	16.45

V. Analysis of Covariance: Criterion - MLA Classroom Writ. Test (LB)[1]

Source	df	Sum Sqs.	Mean Sq.	F
Between	2	861.668	430.834	1.322
Within	22	7171.199	325.963	
Total	24	8032.867	334.703	

Group Means: MLA Writing Test

	Original	Adjusted
TLM	55.22	55.18
FSG	43.41	44.57
FSM	41.23	40.31

---

[1] 10% random sample of each class

ANALYSIS OF COVARIANCE BY LABORATORY SYSTEM

The effects of the language laboratory treatments on the listening and speaking skills were also examined by a straight forward analyses of covariance using the Modern Language Aptitude Test as a pre-measure. The three laboratory systems employed in project schools were confined to the "Functional Skills" strategies. All classes utilized a classroom tape recorder for dialog and pattern practice. This was viewed as the baseline or "control" treatment (TR).

Other classes were assigned to two one-half hour practice sessions per week in either Audio-Active (AA) or Audio-Record (AR) language laboratories in emulation of the prevailing practice in laboratory utilization among secondary schools in Pennsylvania. A survey completed after the close of the experimental instruction (October, 1968) revealed that the twice weekly laboratory usage was still typical of secondary schools.

A summarization of these laboratory system comparisons is reported in Table 28. It can be seen that there were no statistically significant differences.

TABLE 28

ANALYSES OF COVARIANCE BY LABORATORY SYSTEM

for French I and German I Classes

Contrast	Mean Sq.	F(df)
I. FRENCH I - FSG (3 TR, 12 AA, 8 AR)		
1. Variance: Pre-measure, MLAT	11.08 81.17	.137 (2,20)
2. Covariance: MLA Listening Test (LA)	13.26 65.78	.202 (2,19)
3. Covariance: MLA Speaking Test (LA)	3.01 7.56	.397 (2,19)
II. FRENCH I - FSM (3 TR, 15 AA, 7 AR)		
1. Variance: Pre-measure, MLAT	26.36 93.73	.281 (2,22)
2. Covariance: MLA Listening Test (LA)	21.34 80.81	.264 (2,21)
3. Covariance: MLA Speaking Test (LA)	18.00 82.00	.221 (2,21)

TABLE 28

(Cont'd)

Contrast	Mean Sq.	F(df)
III. GERMAN I - FSG (5 TR, 9 AA, 4 AR)		
1. Variance: Pre-measure, MLAT	$\frac{32.43}{117.51}$	.276 (2,15)
2. Covariance: MLA Listening Test (LA)	$\frac{19.97}{168.38}$	.119 (2,14)
3. Covariance: MLA Speaking Test (LA)	$\frac{.84}{5.97}$	.141 (2,14)
IV. GERMAN I - FSM (4 TR, 10 AA, 5 AR)		
1. Variance: Pre-measure, MLAT	$\frac{45.57}{69.90}$	.652 (2,16)
2. Covariance: MLA Listening Test (LA)	$\frac{167.12}{111.87}$	1.494 (2,15)
3. Covariance: MLA Speaking Test (LA)	$\frac{12.67}{8.26}$	1.533 (2,15)

LANGUAGE LABORATORY USAGE

A number of readers of the Research Reports have questioned the employment of the language laboratory in the restricted application permitted by the experiment. Others have been unable, perhaps due to unclear text, to determine exactly how the language laboratory systems were employed (see Appendix B, 5-0683).

Within the framework of the research, three types of audio assistance systems were specified for use by "Functional Skills" classes:

1. A classroom tape recorder (TR) to be used on a daily basis for teacher directed pattern practice drills and pronunciation exercises;

2. Two twenty-five minute periods per week were devoted to class use of either an audio-active (AA) or an audio-record (AR) language laboratory system. (e.g. p. B-7, 5-0683)

In all cases only the commercially prepared audio programs that accompanied the particular text were in use by the class. "Traditional" classes occasionally had a tape recorder for playing music or cultural tapes but not pattern drills even when such tapes had been produced by the publisher.

Why the imposition of twice weekly usage only on participating classes? During the planning stages of the research study it became apparent that twice weekly utilization of laboratory facilities was by far the most frequent pattern among Pennsylvania secondary schools. This pattern apparently had its basis in limitations of space and facilities. A number of schools reported that classes used the language laboratory only once each week. This was increased to make these classes conform to the experimental treatment.

Hayes (1963, p. 20) had pointed out that, "In view of the indispensable requirement of frequent, regular practice, equipment should be provided to allow at least twenty minutes use per class day per student. This means that...it may be advisable to install equipment far simpler than that described in Chapter I /AA and AR laboratories/."

For this reason the Objective 2 of the research was to assess which laboratory system "is best suited economically and educationally..." The original research hypothesis was not, then, which is the ideal language laboratory system and usage combination but an assessment of the language laboratory in the actual school. Was the laboratory being employed by secondary schools in the most economically justifiable manner? This was the purpose of the study.

CONCLUSIONS

An examination of the results of the analyses of covariance based on pre- to post- measures without using midyear tests indicates more significant differences than the analyses reported for USOE Projects 5-0683 and 7-0133. The results of the preceding analyses are summarized for clarity in Table 29.

Significant differences existed in favor of the "Traditional" classes after both Level I and II on both French and German reading measures. "Traditional" classes achieved significantly better than "Functional Skills" classes on French reading and writing tests and as well as "Functional Skills" classes on the listening test. A similar but less significant pattern can be seen for German I and II.

The language laboratory still seems to have had no effect on achievement in either listening or speaking among "Functional Skills" classes.

In summary, an analysis of covariance by class means using the Modern Language Aptitude Test as a covariate indicates more significant achievement for classes using an up-dated cognitive "Traditional" approach to second language learning than previously reported analyses. This trend continues into advanced levels of foreign language study.

TABLE 29

SUMMARY OF REANALYSES
USOE Projects 5-0683 and 7-0133
Analyses of Covariance by Class Means, Tables 23-28

Covariate: Modern Language Aptitude Test (pre-experimental)
Criteria: MLA Cooperative Classroom Tests (post-experimental)

Contrast	Test	French I (10 TLM, 23 FSG, 25 FSM)	German I (6 TLM, 18 FSG, 19 FSM)
I. A. By Strategy	1. Listening	n.s.	n.s.
	2. Speaking[1]	TLM $> p < .05$	n.s.
	3. Reading	TLM $> p < .05$	TLM $> p < .05$
	4. Writing[1]	TLM $> p < .01$	n.s.
		(3 TR, 12 AA, 8 AR)	(5 TR, 9 AA, 4 AR)
B. By System At FSG	1. Listening	n.s.	n.s.
	2. Speaking[1]	n.s.	n.s.
		(3 TR, 15 AA, 7 AR)	(4 TR, 10 AA, 5 AR)
C. By System At FSM	1. Listening	n.s.	n.s.
	2. Speaking[1]	n.s.	n.s.
		French II (4 TLM, 14 FSG, 6 FSM)	German II (6 TLM, 9 FSG, 11 FSM)
II. By Strategy	1. Listening	n.s.	n.s.
	2. Speaking[1]	n.s.	n.s.
	3. Reading	TLM $> p < .05$	n.s.
	4. Writing[1]	TLM $> p < .05$	n.s.

---

[1] 10% random sample of each class

STUDENT OPINION SHIFTS

In USOE Projects 5-0683 and 7-0133 the Student Opinion indexes for various strategy/language/sex combinations had been analyzed by analyses of variance with subsequent Tukey "A" multiple range tests to determine which means contributed to statistical significances. In chosing this procedure, certain analysis and groupings were combined due to the great amount of time required to score and calculate the analysis. An IBM 1620 computer, for example, scored one student every seven seconds, requiring a two hour computer run to make a single check among French classes. Later, installation of an IBM 1401 reduced this same time to an hour--still a prohibitive amount of time.

With an IBM 360 system, it was possible to score and place the 17,000 Student Opinion Scales on tape for fast retrieval. This permitted analyses of covariance on student opinion shifts by strategy. Tables 30 and 31 illustrate French opinion changes among 1,386 Level I and 371 Level II students. In Level I, "Traditional" and "Functional Skills Grammar" students opinion indices dropped significantly more than the pure "Functional Skills Method" students. This was not true during Level II.

German students (N = 1039) did not differ significantly by strategy after Level I (Table 23). Among the 453 Level II students, however, the audiolingual "Functional Skills Method" students indicated significantly lower opinions of foreign language study than their counter parts in other strategies (Table 32; TLM FSM, p .05; FSG FSM, p .01.)

TABLE 30

ANALYSES OF COVARIANCE OF FRENCH I STUDENT OPINION SHIFTS
Student Opinion Scale (1 = low, 7 = high)

Group	N	Pre-Exper. SOS Mean	Post-Mean	Adjusted Mean
TLM	208	5.42	4.80	4.78
FSG	593	5.39	4.87	4.86
FSM	585	5.35	5.00	5.02

Analysis of Variance for Pre-Experimental Opinion Scale (Sept. 1965)

Variation	D/F	Mean Sq.	F
Between	2	.496	.886
Within	1383	.560	
Total	1385	.560	

TABLE 30

(Cont'd)

Analysis of Covariance: Criterion, Final *Opinion Scale* (May, 1966)

Variation	D/F	Mean Sq.	F
Between	2	5.852	5.868**
Within	1382	.997	
Total	1384	1.004	

Finney t-test for differences between means:

TLM-FSG t = .98 at 1382 df
TLM-FSM t = 2.94** at 1382 df
FSG-FSM t = 2.71* at 1382 df

---

*p .05
**p .01

TABLE 31

ANALYSES OF COVARIANCE OF FRENCH II STUDENT OPINION SHIFTS
Student Opinion Scale (1 = low, 7 = high)

Group	N	Pre-Exper. SOS Mean	Post-Mean	Adjusted Mean
TLM	41	5.45	4.61	4.60
FSG	98	5.51	4.76	4.72
FSM	232	5.39	4.87	4.89

Analysis of Variance for Pre-Experimental *Opinion Scale* (Sept., 1965)

Variation	D/F	Mean Sq.	F
Between	2	.531	1.067
Within	368	.498	
Total	370	.498	

Analysis of Covariance: Criterion, Final *Opinion Scale* (May, 1967)

Variation	D/F	Mean Sq.	F
Between	2	1.965	1.913
Within	367	1.027	
Total	369	1.032	

TABLE 32

ANALYSES OF COVARIANCE OF GERMAN I STUDENT OPINION SHIFTS
Student Opinion Scale (1 = low, 7 = high)

Group	N	Pre-Exper. SOS Mean	Post-Mean	Adjusted Mean
TLM	149	5.38	5.09	5.10
FSG	464	5.42	5.03	5.03
FSM	426	5.41	5.03	5.03

Analysis of Variance for Pre-Experimental Opinion Scale (Sept., 1965)

Variation	D/F	Mean Sq.	F
Between	2	.078	.167
Within	1036	.467	
Total	1038	.467	

Analysis of Covariance: Criterion, Final Opinion Scale (May, 1966)

Variation	D/F	Mean Sq.	F
Between	2	.335	.354
Within	1035	.946	
Total	1037	.945	

TABLE 33

ANALYSES OF COVARIANCE OF GERMAN II STUDENT OPINION SHIFTS
Student Opinion Scale (1 = low, 7 = high)

Group	N	Pre-Exper. SOS Mean	Post-Mean	Adjusted Mean
TLM	105	5.35	5.03	5.06
FSG	145	5.53	5.13	5.10
FSM	203	5.41	4.74	4.75

Analysis of Variance for Pre-Experimental Opinion Scale (Sept., 1965)

Variation	D/F	Mean Sq.	F
Between	2	1.076	2.465
Within	450	.437	
Total	452	.439	

TABLE 33

(Cont'd)

Analysis of Covariance: Criterion, Final *Opinion Scale* (May, 1967)

Variation	D/F	Mean Sq.	F
Between	2	6.33	6.092**
Within	449	1.04	
Total	451	1.06	

Finney t-test for differences between means:

TLM-FSG t = .38 at 449 df
TLM-FSM t = 2.37* at 449 df
FSG-FSM t = 2.83** at 449 df

---

*p .05
**p .01

REGRESSION ANALYSES

To more fully assess the influence of various experimental variables on student achievement, regression analyses were computed using student and teacher measures as predictors. Criterion measures were the *MLA Cooperative Classroom Tests*, Level I.

All fifteen predictors were able to account for from 16.5 to 52% of the variance on criterion measures. The greatest influences on variance were language aptitude in French as measured by the *Modern Language Aptitude Test*, and the Language I.Q. score of the *California Test of Mental Maturity* (Short Form).

Teacher experience or graduate training did not seem to contribute greatly to student achievement. Teacher scores on the *MLA Proficiency Test for Teachers and Advanced Students* also did not contribute greatly to student success except in a few cases—French reading and writing (where the contribution is negative); German listening, speaking, and writing (contribution positive). In general, teachers scores on the *Writing Proficiency Test* seemed to influence student achievement more than any other measure.

TABLE 34

REGRESSION ANALYSIS: FRENCH I FINAL ACHIEVEMENT
178 student random sample of 57 classes

Predictive Variables Pre-experimental	Criterion Variables, Final MLA Cooperative Classroom Tests LA Listening Beta	 LA Listening %Variance	 LA Speaking Beta	 LA Speaking %Variance	 LA Reading Beta	 LA Reading %Variance	 LA Writing Beta	 LA Writing %Variance
1. Lang. IQ	-.134	-1.2	.035	1.0	.105	3.2	.063	1.7
2. Mod. Lang. Apt. Test	.402	15.3	.391	15.2	.354	15.5	.446	19.3
3. LA List. Test	.244	7.2	.220	7.0	.273	9.2	.046	.8
4. Tchr. grad. hrs.	-.106	.4	-.073	.2	.136	1.0	.054	.3
5. Tchr. yrs. exper.	.090	.6	.182	1.9	.074	1.0	.115	2.1
6. Tchr. self est.: Spk.	.162	.3	.172	.6	.218	1.9	.197	-.3
7. Tchr. self est.: Read.	.023	--	-.107	.4	-.186	.9	-.149	1.6
8. Tchr. self est.: Write.	-.113	.5	.037	--	.020	--	-.023	.1
9. Tchr. Prof.: Spk.	-.155	-1.2	-.222	2.4	.092	.1	-.106	1.6
10. Tchr. Prof.: List.	.062	.5	.439	1.6	.129	-.2	.197	-2.3
11. Tchr. Prof.: Read.	-.113	-.7	-.388	6.2	-.323	2.4	-.366	7.4
12. Tchr. Prof.: Write.	.012	.1	.292	-1.1	.252	-.2	.151	-2.0
13. Tchr. Prof.: Ling.	.133	1.8	-.183	2.4	-.130	1.3	.064	-.9
14. Tchr. Prof.: Cult.	.222	4.3	-.002	--	.045	--	.081	-.5
15. Tchr. Prof.: Prep.	-.058	-.3	-.085	.8	-.189	2.0	-.339	8.6
%Variance accounted for		27.6%		38.2%		38.1%		37.5%
Coefficient Mult. Regression		R = .53**		R = .62**		R = .62**		R = .61**
Mult. Regression Equation		[.14 V2 + .46 V3 + .18 V14-4.83] R = .48**		[.25 V2 + .64 V3 + 25 V5 + .54 V10 - 53 V11-.28 V15 + 9.61] R = .58**		[.07 V1 + .12 V2 + .34 V3-11 V15 + 5.39] R = .55**		[.48 V2 - V + .49 V14 - .64 V15 + 37.96] R = .57**

** $p < .01$

TABLE 35

REGRESSION ANALYSIS: GERMAN I FINAL ACHIEVEMENT
111 student random sample of 41 classes

Predictive Variables Pre-experimental:	Criterion Variables, Final MLA Cooperative Classroom Tests							
	LA Listening		LA Speaking		LA Reading		LA Writing	
	Beta	%Variance	Beta	%Variance	Beta	%Variance	Beta	%Variance
1. Lang. IQ	.253	8.6	.244	9.0	.050	.5	.186	7.2
2. Mod. Lang. Apt. Test	.136	3.8	.255	11.5	.175	4.6	.471	26.2
3. LA List Test	.224	4.0	.163	2.0	-.068	.4	.133	1.5
4. Tchr. Grad. hrs.	.031	-.2	-.027	.2	.184	2.7	.104	-.1
5. Tchr. yrs. exper.	.045	.2	.074	1.3	-.081	-.5	-.154	-.1
6. Tchr. self est: Spk.	-.319	4.7	-.406	9.9	-.058	.2	-.356	5.0
7. Tchr. self est: Read.	.201	-2.9	.088	-2.0	-.394	3.0	.167	-1.7
8. Tchr. self est: Write.	-.074	.7	.062	-.9	.230	-.4	-.024	.2
9. Tchr. Prof.: Spk.	.055	.8	.225	7.0	.122	2.3	-.232	-1.7
10. Tchr. Prof.: List.	.183	3.6	.056	1.6	.075	.9	.047	.6
11. Tchr. Prof.: Read.	-.278	-1.7	-.121	-2.7	.064	.5	-.038	-.2
12. Tchr. Prof.: Write.	.446	6.4	.412	12.6	-.188	-2.1	.711	12.7
13. Tchr. Prof.: Ling.	-.370	-1.0	-.186	-4.1	.181	2.2	-.289	-2.8
14. Tchr. Prof.: Cult.	.013	.04	-.197	-.3	-.238	1.6	-.296	3.9
15. Tchr. Prof.: Prep.	.15	1.9	.120	2.7	.045	.6	.085	1.3
%Variance accounted for		29.3%		47.8%		16.5%		52.0%
Coefficient Mult. Regression		R = .54**		R = .69**		R = .41*		R = .72**
Mult. Regression Equation	[.17 V1 + .31 V3 -.92 V6 + .09 V10 -13.02] R = .45**		[.21 V1 + .17 V2 -2.58 V6 + .37 V9 -17.88] R = .63**		[.08 V2 + .17 V9 = .12 V14 + 11.3] R = .34**		[.29 V1 + .42 V2 -24.0] R = .59**	

*p <.05
**p <.01

TABLE 36

REGRESSION ANALYSES
STUDENT AND TEACHER EFFECTS ON MEAN CLASS ACHIEVEMENT, LEVEL I

Criterion: Foreign Language Achievement (MLA Cooperative Classroom Tests, converted L+S+R+W Scores)

Prediction Variables	FRENCH (N = 57) Mean	S.D.	r	Beta	% Variance	GERMAN (N = 41) Mean	S.D.	r	Beta	% Variance
1. Class size	21.28	5.27	.15	-.096	1.39	21.95	5.80	.25	.068	1.66
2. Class Lang. I.Q.	110.87	16.02	.25	.115	2.84	114.30	4.47	.41	.392	16.23
3. Class MLAT	46.23	8.32	.41	.442	18.02	46.18	9.06	.11	.047	.54
4. Class pre-exp. LA List.	9.58	1.60	.00	-.070	-.03	10.80	1.81	-.02	.049	-.08
5. Tchr. Grad. hrs.	36.32	39.80	.05	-.083	-.38	45.80	44.47	-.01	.204	-.18
6. Tchr. Yrs. Exper.	9.56	8.91	.17	.238	3.93	9.54	9.49	.08	.019	.16
7. Tchr. Self-Spk.	1.70	1.02	-.11	-.090	1.01	1.51	1.14	-.29	-.101	2.92
8. Tchr. Self-Read.	2.02	1.04	-.12	-.143	1.71	2.07	1.19	-.26	.233	-6.07
9. Tchr. Self-Writ.	1.67	1.02	-.09	.086	-.78	1.68	1.08	-.28	-.624	17.68
MLA Proficiency Tests										
10. MLA Prof. Listen.	39.16	7.90	-.03	-.026	.08	41.00	8.13	.00	.044	.00
11. MLA Prof. Speak.	74.33	19.69	-.03	-.112	.31	87.10	13.94	.01	-.120	-.08
12. MLA Prof. Read.	45.68	10.46	-.05	-.374	1.84	50.05	10.30	.05	-.252	-1.13
13. MLA Prof. Writ.	45.46	9.95	-.01	.312	.37	55.15	12.89	.13	.601	7.76
14. MLA Prof. Ap. Ling.	49.26	9.38	.08	.037	.30	52.37	7.59	.10	.102	-1.04
15. MLA Prof. Cult. & Civ.	47.26	9.58	.04	.202	.77	51.49	8.62	-.06	-.040	.25
16. MLA Prof. Prep.	63.32	10.86	-.103	-.244	.64	62.85	6.53	.02	.079	.12

Constant term -5.23 -49.05
Coefficient of Multiple Regression .57 (df 1,56) .62 (df 1,40)
Coefficient of Multiple Determination .32 .39

REACTIONS AND REVIEWS OF THE RESEARCH

SECTION IV

The reactions of the profession to the findings of the research were slow in starting, perhaps reflecting the summertime distribution of the Final Report of Project 5-0683. The report was formally submitted to the USOE in March, 1968. In June and July, after notification of USOE acceptance, several hundred copies of the report were mailed to the state supervisors and leading foreign language educators throughout the nation. Two months later, in mid-September, West Chester State College released the results of the study to the public.

First professional reporting were the Bulletin of the Pennsylvania State Modern Language Association (October), the Ontario Educational Review (November) and in Lingua, the Swedish Modern Language Journal. Subsequently, the reports have been mentioned in a wide variety of media from syndicated newspaper columns to Education Today. The study will be discussed in detail in the October, 1969, Modern Language Journal and the December, 1969, edition of Foreign Language Annals.

Selected comments on the results of the research project to date include:

> "(The City Supervisor) is hiding your report" --Professor, a Pennsylvania university
>
> "... very dangerous" --City Supervisor, Pennsylvania
>
> "... compares well with the Keating Report." (comment at MLA)
>
> "Many of us only hope that Pennsylvania will not go backwards despite the findings of your research." --University of Massachusetts
>
> "... We are eagerly looking forward to your follow-up study" --University of Goteborg, Sweden
>
> "I admire you for courageously stating conclusions and implications even though they will make some people in the field very unhappy." Junior College President
>
> "... our congratulations and our admiration" --Dept. Linguistics, University of Edinburgh
>
> "... a milestone in the history of methods of teaching foreign languages not only in this country but also in the rest of the civilized world." --Chairman of a Language Department, State University of New York

There can be no doubt that the findings of the project are either most encouraging or disturbing, depending upon the biases and receptivity of the reader. Certainly, the conclusions were personally traumatic to the project staff, deeply committed to a "Functional Skills" philosophy.

There can also be no doubt that in a study the size and scope of the Pennsylvania project there are bound to be errors--errors in planning, analyzing, reporting, duplicating and interpretation. Some of these are obvious with the brilliant illumination of hindsight. Others are more technical, depending upon basic assumptions of statistical precedures. Some are simply oversights due to the enormity of the study. A few depend upon viewing the project as it was intended, a curriculum assessment of already implemented innovation rather than an "original" research study.

One general criticism of program evaluation, prematurity, is not appropriate. The study was not implemented until after the "Functional Skills" approach had become widely accepted in both professional thinking and actual school implementation. Even in such a small, traditionally conservative, rural state as Nevada, for example, ninety-five per cent of the secondary schools had adopted the "Functional Skills" approach by the 1963-64 school year.

This portion of the SUPPLEMENTARY REPORT has two primary objectives: (1) to provide readers with the observations of extra-project professionals on the research study and (2) to provide additional information, analyses, and comments on the first two years of instruction.

The first objective is achieved by reproducing available formal reactions to the Pennsylvania Foreign Language Research Project. Reviews include the proceedings of a formal discussion conference held in March, 1969, on the West Chester State College campus which brought together again as many of the original project planners and consultants as possible. Three of the consultant panel could not participate, one due to health and one due to a sabattical leave. The third consultant did not acknowledge receipt of several communications from the project inviting his participation. This transcript has been edited for clarity and annotated on occassion.

In addition, comments by Albert Valdman, Rebecca Valette and Kenneth Lester are included. These are reproduced exactly as submitted by the authors with no changes, annotations, or additions.

DISCUSSION CONFERENCE ON USOE PROJECTS 5-0683 AND 7-0133, WEST CHESTER, PA., MARCH 20, 1969

PREFACE

Since 1965 the Pennsylvania Foreign Language Project, a joint effort of the Bureau of Research of the Pennsylvania Department of Education and West Chester State College, has been conducting a research [illegible] selected schools throughout the state.

Smith, P. D. and Berger, E. *An Assessment of Three Foreign Language Teaching Strategies Utilizing Three Language Laboratory Systems.* Final Report of USOE Project 5-0683, January, 1968. ERIC Document ED 021 512.

Smith, P. D. and Baranyi, H. A. *A Comparison Study of the Effectiveness of the Traditional and Audiolingual Approaches to Foreign Language Instruction Utilizing Laboratory Equipment.* Final Report of USOE Project 7-0133, October, 1968.

On March 20, 1969, the project conducted a discussion conference on the research study and the reports in order to provide the profession with a critical review of the assessment.

This document is a condensation of the discussion meeting, abridged to avoid lengthy introductory remarks, edited for clarity and relevance, and provided with notes where necessary.

Participants in the discussion conference included:

Helmut Baranyi	University of Pittsburgh Project Staff
Emanuel Berger	Pennsylvania Department of Education Principal Investigator
John Carroll	Educational Testing Service American Council on Teaching of Foreign Languages
John Crew	West Chester State College Associate Director of Research
Chauncy Dayton	University of Maryland Consultant, Statistical Analysis
Ralph Eisenstadt	West Chester State College Project Staff
Carl Epstein	United States Office of Education Project Officer
Robert Hayes	Pennsylvania Department of Education Director, Bureau of Research Administration
Martin Higgins	West Chester State College Director of Research
LaMarr Kopp	Pennsylvania State University Associate Dean Liberal Arts American Association of Teachers of German

Robert Lado	Georgetown University Dean, Institute Languages and Linguistics
Willard Martin	Pennsylvania State University National Association Language Lab Directors
Julia Petrov	United States Office of Education Institute for International Studies
Alfred Roberts	West Chester State College Project Staff Chairman Department of Foreign Languages
Philip Smith	West Chester State College Project Coordinator
Wilmarth Starr	New York University, Department of French
Albert Valdman	Indiana University Chairman, Department of Linguistics
Milton Woodlen	Eastern Regional Institute for Education Former Project Coordinator
Genelle Caldwell	Delaware State Department of Education Foreign Language Consultant
David Chestnut	Pennsylvania Department of Education Foreign Language Specialist
Peter Esseff	United States Office of Education Higher Education
Paul Glaude	State University of New York Foreign Language Supervisor
Paul Hilaire	New Jersey State Department of Public Instruction Foreign Language Specialist
Roy Hinchelwood	New York University American Association of Teachers of French
Everett Landin	West Chester State College Director, Educational Development Center
Martin Yanis	Pennsylvania Department of Education Bureau of Research

OBSERVERS

Fred Zimmerman	Lock Haven State College
Irene Kent	Lock Haven State College
Nora Huergo	Lock Haven State College
Elizabeth Newton	Kutztown State College
Henry Christman	Kutztown State College
Edward Dulak	Mansfield State College
Ruth J. Kilchenmann	Shippensburg State College
Patricia Annable	Slippery Rock State College
Blossom Brooks	East Stroudsburg State College
Arthur Arnold	East Stroudsburg State College

CONDENSATION OF DISCUSSION CONFERENCE PROCEEDINGS

The conference was opened by Dr. Alfred Roberts who welcomed the participants. Dr. Earl F. Sykes, President of West Chester State College, extended the official greetings of the college. Dr. Sykes pointed out the far reaching impact of the Foreign Language Project and that "... there are going to be many misinterpretations as well as constructive interpretations." He concluded with the comment that progress comes through upsetting the equilibrium and that the research has accomplished a great deal in that it will force rethinking on theories, concepts and approaches to foreign language teaching.

Philip Smith, project coordinator, reviewed the history of the project and the reasons for holding a conference to discuss the research and its implications. Smith indicated that a number of readers of the research reports questioned the research design, control of experimental variable and the statistical treatment. Smith asked for reactions of the group to the appropriateness of the basic research design, the Campbell and Stanley #10, "Non-Equivalent Control Group." Was it a wise choice?

Berger pointed out that the research originally hoped to support the Department of Education in its push for foreign languages. The original proposal justified the type of design. First, the other possibility, the Campbell-Stanley design #4, the purely experimental design which calls for random assignment of students was just not feasible for a state agency. It has no control over local conditions. The state cannot go into a classroom and say "Would you mind giving us a roster of your kids. We would like to randomly assign them. We have some ideas on how we would like to do an experiment?" That type of research could not be done at the state level.

Woodlen, who made the assignments, said that although teachers believed they were given their first or second choice, there was truly a random assignment to treatments.

Higgins asked where the experimental population was obtained.

Berger stated that, first, all school districts were surveyed as to which had and which did not have language laboratories. Also all school districts surveyed identified teachers whose major responsibility was the teaching of French or German. At least a full year of prior teaching experience was required.

Annable asked if there was a check on the ability or the background of the teacher involved with something as specific as audiolingual approach.

Berger commented that there was a check, but not a selection. Remembering that a state agency doing research wants to generalize to local school districts as practiced. The criticism normally leveled at this kind of research is answered by the full week in-service session a week or two just before the beginning of the school year. The classes were a sample of what goes on generally in Pennsylvania.

Smith pointed out that forty of the hundred had been through the NDEA Institute Programs, twice the state average.

Berger reminded the group that, independent of descriptive experience, the teachers were given the MLA Teacher Proficiency Tests as required in Pennsylvania. The measures indicate that they all scored above the passing on all measures.

Carroll thought that the problem of the sample and the generalization of the population is solely one of whether there was any sampling bias which would interact with the variable under study. Of course, this is very difficult to tell. Normally in a study of this sort, with random assignment of classes and a study of certain treatments it would not make any difference as long as there is a reasonably good sample.

Carroll also remarked that the reports questioned the ETS norms on the MLA Cooperative Classroom Tests. In the norms booklet there are lists of schools that were included in the norming population. He wondered whether these were schools that were excluded from the study. Could the project staff make any comment about the kind of sample drawn from Pennsylvania versus the kind of sample ETS had.

Smith pointed out that both ETS and the Pennsylvania project had used some of the same schools. The ETS list contains schools which are not typical. Schools listed in the ETS norming population for other states included in the audiolingual classes some known from personal observation that were not audiolingual in 1963. ETS took the word of the teacher. The project sample from Pennsylvania equaled the total national sample of ETS. The samples were not the same. Many schools in the ETS sample were private schools. The project sample only included public schools.

Woodlen remembered that at the time the study was undertaken ETS was very much interested in the work because the ETS tests had not yet been validated on the same group of students across the board. ETS norms were part norms and were based on a separate sampling group for each segment of the test.

Smith stated that he had contacted ETS and offered them the project data and was discouraged.

Woodlen observed that this was interesting, because ETS gave the project an advantage on the price of the tests in order to get these data.

Berger commented that the problem that has bothered him personally was that the "Traditional" group was not totally randomly assigned. These were people who were already teaching traditionally.

Starr thought that one of the possible areas of muddiness in the design was in not clearly discriminating between the "Traditional" group and the other groups. The list of text books used in French would not be "Traditional" text books in his opinion.

Starr was concerned about the effect on "Traditional" teachers of the orientation meeting and what the effect on the "Traditional" students would be of being exposed to listening comprehension tests and other devices which are characteristic of the "Audiolingual" method. He was concerned if the study really tested or researched what is stated in the design as being researched.

Smith asked for a definitive statement on the appropriateness of the research design from the group.

Carroll believed that, in the abstract, the design was fine. Starr agreed.

Smith reviewed the language laboratory treatments. "Traditional" classes were allowed to have in their room a tape recorder or record player but were not allowed to play tapes containing pattern drills or similar materials. Songs and cultural items were permitted but teachers were not allowed to play laboratory type drills or dialogues.

Tape record classes in "Functional Skills" cells had a tape recorder which they were to use about ten minutes a day in the classroom to play and practice pattern drills. Laboratory classes also had a tape recorder for about 10 minutes daily and in addition they had two half-hour periods per week in the language laboratory.

Carroll questioned homework assignments for project students.

Eisenstadt stated there were no controls on homework since the study was involved with the intact classroom. The study could not effectively limit or increase homework assignments from what school policy normally was.

Carroll felt that homework might have made some difference. That is, it could be argued that under a traditional regimen of instruction there might be more pressure on the students to have homework assignments and to do them, whereas, in the "Audiolingual" technique the student is very often told that the most important learning is going to happen in the classroom.

Roberts pointed out that the answer to Carroll's observation was that "If this is what is done in traditional teaching, that is what we wanted done." Starr earlier raised the question, "Were traditional teachers contaminated by the workshop orientation?" Roberts felt that "Traditional" teachers were contaminated perhaps to the extent that the profession was giving lip service by that time to the "Audiolingual" approach. The only complaint on assignment was from teachers who were assigned to the "Pure Audiolingual" approach. Most of the teachers wanted to give grammar.

Valdman observed that the chief variables could have been contaminated by this bias.

Starr asked whether there really was a significant difference in the strategies? The nucleus of grammar teaching in the TLM and FSG would probably tend to erase any discrimination.

Lado had a basic objection to the logical reasoning that follow from large experimental designs. It came out in the Chicago investigation, it seems to be coming out again. Lado did not object to the particular design, but the argument that by having a large study with a lot of schools that one is going to learn something realistic. The argument is that by involving many schools one finds which of the methods is the best. Lado felt that it is inherently impossible to have anywhere near satisfying controls in such a vast undertaking. For example, had the researchers noted if any of the students went to France or Germany during the summers. The study did not indicate whether the family backgrounds of students were French or German. There was, better than other experiments, initial testing of proficiency. However, it is practically impossible to control any mass experiment like this to a point to where one can be really satisfied about it. Therefore, one cannot assume that if "Traditional" classes achieved better results it is due to the method.

Lado favored working on realistic classroom situations that can be controlled. If one then finds something from the controllable situation, if results show a difference, one can ask "Why?" and look for more productive answers.

Higgins remarked that the large group type of design has a lot of limitations. However, one of its advantages was referred to in Lado's statement. There can be atypical students in the data--those who go to Europe, people who have bi-lingual homes. With large numbers in each treatment one can assume with a fair degree of certainty that these would be evened out.

Lado replies that he realized this is the argument, but that this failed to satisfy him. Lado also felt that the tests used to measure achievement were not enough. A twenty or thirty minute test of listening comprehension is not sufficient. Given the size of the Pennsylvania experiment, one could not do much better. It is a remarkably, carefully thought out, design or experiment from the point of view of trying to include tests. If it had been in a more manageable dimension, there could have been a number of smaller comparisons.

Esseff reminded the group that professionals involved in technology have attempted comparative studies in closed-circuit television, programmed instruction, and computer assisted instruction and have come up with no significant differences. Professionals in the field of educational technology are disenchanted with the comparative approach and feel that the many complex variables that are inherent in the media do not lend themselves to comparative analyses.

Hayes reminded Esseff that a state agency was interested in down-to-earth research with practical implications to find out whether or not new technology is really worth the expense involved. Are new methods really better than the old ones? Can these be determined without a comparative approach.

Starr asked again if the project staff really knew that the methods were different.

Higgins pointed out that this is the big weakness of large scale research--adherence to treatment. How does one know if the teacher once assigned a condition behaved as she was recorded, and did Condition X differ from Condition Y.

Glaude felt that the test materials used were not unalloyed as a "Traditional" text. The project may compare transitional materials and transitional teachers. The borders are just not clear.

Woodlen spoke on the selection of schools and materials based upon preliminary surveys of state department information and project questionnaires concerning texts in use in participating schools. This information was presented to the panel of experts in a two-day session. The staff listed twenty-seven text books in use in French and twenty-eight in German. The panel then decided, after reviewing all of the texts, on three "official" "Traditional" text books or materials and either the A-LM or the Holt Rinehart materials for "Functional Skills" classes. Schools that did not have these materials were aided by project funds in obtaining proper texts. Research in public school situations encounters real restraints on testing time. The first plan included five days of pretesting. This was reduced to three and a half but even this caused some teachers unhappiness.

Smith observed that the text book has a built-in emphasis in theory. The texts, according to Woodlen, were selected by the panel of consultants.

Starr did not remember selecting text books but believed that they are all, more or less, "Audiolingual," even the "Traditional" texts.

Nota Bene: The texts books used in the research and under discussion at this point were:

FRENCH
Traditional: *Cours Elementaire de Francais*
Dale and Dale, 1st ed., 1949,
2nd ed., 1956.
Parlez-Vous Francais?
Huebener and Neuschatz
2nd ed., 1958.
New First Year French
O'Brien and LaFrance
1st ed., 1958.
Functional Skills: *Audio-Lingual Materials*
1st ed., 1961
Ecouter et Parler
Cote, Levy and O'Conner
1st ed., 1962.

GERMAN
Traditional: *A First Course in German*
Huebener and Newmark
2nd ed., 1964.
Foundation Course in German
Homberger and Ebelke
Rev. ed., 1964.
Functional Skills: *Audio-Lingual Materials*
1st ed., 1961
Verstehen und Sprechen
Rehder, Twaddell and O'Conner
1st ed., 1963.

Valdman believed the most vulnerable part or aspect of the project is the definition of the three strategies. Secondly, the control and the implementation. The consultants were concerned about contamination--concerned whether, in fact, it was possible to define strategies.

Lado stressed that consultants suggested ways in which the research could be improved but did not design the study from the beginning.

Valdman remembered that the consultants' first question was whether the project planners really desired to have three different strategies rather than two.

NOTE: In the planning stage a fourth strategy had been suggested but not included (Report 7-0133, p. 21).

Berger said that a great deal of thought had been given to all the controls, not just the book but all. The text is a very significant factor. Certainly, the responsibility for this type of research project is upon those who conducted it and wrote it. The consultants bear no responsibility for suggestions accepted or rejected. This should be very clear. The texts are those the profession identified.

What is critical is the whole technique of teaching. In the "Traditional" text there is a very heavy emphasis on a presentation of grammar exercises and vocabulary control. There may be other techniques suggested by newer ideas that the authors introduced because they were convinced that this was the thing to do. But the text obviously determined the method.

The A-LM and the Holt-Rinehart materials were considered by the professionals to represent drastically different philosophies of teaching. So if the "Traditional" have moved somewhat, it was still fairly consistent from the standpoint of the underlying philosophy of the way languages should be taught. The role of grammar, paradigms, vocabulary, idiomatic expressions, culture (which is really incidential) are shaped by the materials. Materials do control and distinguish between the two strategies.

Roberts added to Berger's observation that Dale and Dale first and second editions, were never referred to as "Audiolingual Texts" but the A-LM and Holt-Rinehart materials are consistently referred to as "Audiolingual."

Martin asked if there should be any speaking expected in the "Traditional" classes.

Smith thought it was a tragic mistake to call the conceptional approach "Traditional" because "Traditional" reflects a time lag. What was "Traditional" to the new breed of state supervisors, 1960-61, was 1955. To others, "Traditional" is 1925. It is what one is not doing himself that is "Traditional." It would have been better to have used terms like "deductive" or Carroll's "Cognitive Code-Learning."

The word "Traditional" has upset a great deal of people, making for bad publicity and bad press. The terms "Audiolingual" and "Grammar-Translation" or "Structured Approach" may be more appropriate.

In response to Martin's question, "Is the teacher ever supposed to speak the foreign language?" Smith observed that the professional has a bad picture of the "Traditional" teacher. Most professionals in language education today learned from very good "Traditional" teachers, some of whom used the language to a considerable degree. Good "Traditional" teachers never divorced the language from the classroom.

The question of how much English was heard should be answered by one of the observers.

Eisenstadt reported in the classroom he observed, treatments were adhered to rather rigorously. In "Traditional" classes the foreign language was not heard very much.

Starr wondered why the expected goals that are listed for the "Traditional Method" under speaking are not significantly different from those listed for the "Functional Skills Method." If students use similar materials, A-LM or Dale and Dale, in both groups; if expected goals includes speaking; and if there was use of the target language in the classroom--then, of course, the results are going to come out as they did.

Starr felt that "Traditional" should have used materials from the 1930's and 1940's with no use of the language and no expected ability to read after the model sounds, words and phrases. Nor should students have the ability to very basic structural patterns by responding to simple questions. That is an "Audiolingual" technique.

Smith disagreed that these were solely expectations of "Audiolingual" techniques.

Higgins believed the acceptance of common objectives certainly does not preclude independence of treatment. One can try to achieve the same objectives applying rather different methods.

Starr then referred to Lado's point of mass versus a small group. If one had a hundred students and did not use the foreign language in the classroom, did not have a tape recorder, did not give them listening comprehension tests, but gave them an old book and taught them in English, basically grammar and written exercises and then compared them to a class which did use the foreign language, had a laboratory and did listening exercises, it would not be a measure of what happens in the mass of Pennsylvania classrooms but it would have been something to talk about. Starr reiterated that he was not shocked by the results because they were predictable. The controls, the design and the descriptions seemed to be collaborating.

NOTE: Descriptions of the general criteria and definitions were the subject of discussion later and have been reproduced earlier in the report. (pp. 6, 7 and 8).

Valdman thought that perhaps a flaw in the experiment was that the rating scales used to control teacher adherence to treatment strategy were not parallel. In the rating scale for adherence to "Traditional" there is no measure of the amount of foreign language used in classroom by either the teachers or the student but this is measured for both "Functional Skills" strategies. It seems it would have been more helpful if the observer could have made some quantitative, even perhaps qualitative, observation of the use of the target language and the native language in various strategies. The rating scale would have set up on a quantitative basis with each type of activity assigned a value from 1 to 5. He thought this would have been the best way to check on teacher adherence.

Smith said the first observation report was replaced at midyear because it was not satisfactory. The original ones were parallel and as he understood it, did not work. The observers could not make these things come out. They were too subjective. Observers were trying, they thought, to make too subjective judgments.

Valdman stated the rating scales made provisions for observation of vocabulary drill. It did not specify the type of vocabulary drill. There is a control for translation of reading lessons; for formal discussions of grammar; that is something which surprised him. There is a control for pronunciation.

Smith corrected Valdman's conception of the Teacher Observation Scales as "Controls." They were not controls but observation.

NOTE: The scales were rated 1-5 as Valdman had suggested they might have been

Valdman stated that this is where the staff controlled the teachers adherence to strategies. They were actually controls and the observation reports may not cover enough items.

Caldwell asked about the drop-out rate in the "Traditional" and "Audiolingual" approaches. Those who work in schools think of "Audiolingual" programs as at least four year programs and not in terms of a one year program as opposed to a two year program. In the report on the third year there were not enough students remaining in the "Traditional" program to have a meaningful comparison with "Functional Skills" students. Was this lack of students significant? One of our accomplishments has been to develop foreign language programs the students can cope with and hopefully over a very long period of time.

Smith reported that there was no valid data on drop-outs. If students lost data they were dropped from the experiment. He was not excluded from the class but from the population. Absentees were not allowed to make up tests. The decision to stay for second and third year was often not a function of the student. The teacher moved or quit. School districts felt that the study had been testing students too much. Drop-outs cannot be studied with the data available.

Caldwell thought it was too bad that the "Traditional" teacher is not described in the way in which she apparently functioned—a person who does use the language in the classroom.

Crew asked if the level or extension of behavior over five hours of testing would have given the researchers more information. This might be assumed or it might not. Assuming one had good testing, the students would have still ranked themselves over twelve hours of testing, in many behaviors, as they would have over thirty or forty hours. This is fair assumption which usually stands if one has good measures in the first place.

Crew's second comment concerned the matter of one hundred students, precise control, and the question of research--the purpose of doing experiments in the first place. If one has twenty-five to fifty students, he has only zeroed in one school, one classroom. This research was started to get some broader idea as to how these things work.

It has also been mentioned that the "Traditional" strategy may have been contaminated. There must have been *some* obvious difference in each of the three groups. If there are not any differences across the broad scope of classrooms, different teachers, different pupils, if one finds the same thing in five states or a thousand schools one could be in a position to say it looks as though none of these may be crucial elements in teaching foreign languages. There must be a "comparison" to ask a basic question in modern languages or any other field.

Smith stated that the profession has done, in effect, some small scale research. The attempt to replicate small studies on a bigger scale does not come out the same.

Higgins' reminded the group that on criterion measures it was necessary to differentiate between reliability and validity. The study deals with class means, a highly stable measure. From that point of view the reliability of the comparison is good. The validity of the criterion measures [*MLA Cooperative Classroom Tests*] were unknown to him.

Higgins second comment concerned adherence to treatment in the three strategies. Researchers acknowledge that there is considerable variability in the extent to which a person adheres to treatment or does not adhere. If one can make the assumption that the deviation from treatment is no more systematic in any of the three treatments then the others, one can still compare treatment effects.

Baranyi pointed out that the project staff did ask teachers at the meeting of May, 1968 whether they stayed within the realm of their teaching outline. Of the fifty some people that were there, more than half of the project teachers, maintained that they were professional enough to stay within their assignment.

Higgins asked whether the students were questioned concerning the kinds of behavior the teachers had used during the year.

Smith stated that this had not been done.

Kilchenmann asked how many teachers were re-educated through NDEA Institutes since she did not think institutes are as significant as supposed. Were there any teachers in the project that had been trained themselves audiolingually from the beginning, from college on, and did not have to be re-educated? There are many so called "Audiolingual" teachers using instead an eclectic method. She felt it important to know how many teachers were superficially audiolinguists and how many were real audiolinguists.

Smith mentioned that precise teacher observational techniques such as the Flander's Interaction Analysis had not been successfully applied to foreign language teaching when the project was initiated.

Glaude agreed that this is a very significant project. The outcomes are very significant. Certainly, if nothing else, it has proven that regardless of the materials or the techniques used there must be a very strong structural focus on outcomes and on instruction at the time it takes place, regardless of any direction one goes. This is good. However, some outstanding techniques have existed a long time but the report does not specifically mention them. Did one see much of that" Did the observers see this" Was this typical? This would be helpful to the reader.

Smith stated that the researchers assumed all the way through, that every teacher has individual techniques that they are going to use within the framework of the very precise teacher's manual.

The test /MLA Cooperative Classroom Tests/used were the best ones that were available, brand new at that time. To test "Traditional" students and behaviors fairly, the project reprinted the 1939-40 Cooperative French and German Tests. Rebecca Valette was employed to develop other tests for us, the Listening Discrimination Test and the Sound Production Test. Psychometric analysis at Penn State indicated that these tests showed promise but were not good enough for use as criteria. The Valette Tests are reported but no conclusions are based upon them.

Lado stated it was clear to him when the experiment started that the measuring instruments were not adequate. Lado said he pleaded that at least there should be something specifically on pronunciation. The reports note that there seems to be some significant difference when it comes to the Valette Test but these have not been validated. Now that is about the only place where the research begins to pin-point.

Smith reminded the group that in the MLA Cooperative Classroom Tests there is a Speaking Test which measures pronunciation and production. The "Traditional" students scored significantly higher on the "Critical Sounds" part of this test. (Report 5-0683, pp. 39, D-24)

Valdman asked if the MLA Speaking Tests were scored at the Educational Testing Service or scored locally.

Smith stated that they were scored at West Chester but that the scorers had been trained at the Educational Testing Service. In addition, people from ETS came to West Chester and observed project scorers while they were scoring tests here. They are in effect ETS scorers, trained by and for ETS.

Baranyi observed that when he was first retained by the project it was as a tester for the students at the middle of the second year of the project in the Pittsburgh area. One of his observations was that the "Traditionally" oriented students tended to say more than did the "Audiolingual", especially on the sections where the student is asked to describe and talk about pictures. The "Traditional" student had more vocabulary. There were a lot of silences when testing "Audiolingual" students who were often frustrated at the end of a year and a half of foreign language.

Berger added that it should be made clear that these tests are not intended for the end of one year of study. At the end of one year of study the student does not have the vocabulary. Valette also makes this point. The Educational Testing Service advised the project that it would have different norms but that this should be unimportant as long as one dealt within the same sample. It was considered appropriate for the project to compare "Traditional" students with their "Audiolingual" colleagues at the end of one year, at the end of a year and a half and at the end of two years to see how well they achieve. The "Traditional" students may have had vocabulary equivalent to the "Audiolingual" group at the end of a year and a half. That may have a very strong bearing not only on speaking but may become obvious in the reading, the listening comprehension and writing as well. Certainly, the first year study has to be looked at that way.

Carroll observed that he had always believed this.

Eisenstadt, one of the scorers trained by Educational Testing Service, agreed that the Speaking Tests in particular support what Valette states in her comments that vocabulary-wise after half a year or one year there was not much vocabulary exposure in the "Audiolingual" approaches. It was a rare instance that one found satisfactory responses in the Speaking Tests in our first year classes. This also involved physiologically frustrating factors to be sure, both on the part of the student and on the part of the scorer.

Carroll wished to make a point about the MLA Cooperative Tests that he has made many times before. He agreed that the tests are loaded with vocabulary but that they do not test other aspects. They do not test the grammar and the other things that are supposed to come out of some of the newer methods. The tests don't bring these out, except in very small measure. The Writing Test would bring other skills out probably more than anything else.

Valdman felt there are many other measurable components that contribute to listening comprehension and speaking skills that the MLA Cooperative Tests do not measure; rapidity of response, rhythm and the speed of response on the part of the student. There are many factors which may be significant in establishing speaking comprehension. The problem is that the profession does not know what contributes to listening comprehension ability nor to speaking. This is an inherent weakness in these tests.

He pointed out that in a comparison experiment (it wasn't really a comparison experiment because the comparison turned out to be incidental) he used the ETS battery both the lower level and the higher level, but in addition we had people from FSI administer their interview. He thought that perhaps it would have been interesting to try out, at least at the third year level, the Foreign Service Institute interview technique. It is a test which deals with the total communication situation in which one needs first to understand and then to respond.

The trouble with the Listening Comprehension Test and the Speaking Test is that these are quite artificial. They do not really reflect the natural communication situation. The Listening Comprehension Test is an unpure test because the student has to read the answers. At the first year level this would clearly be a serious disadvantage for "Functional Skills" students who can not read and do not have the vocabulary.

Why is it (except for economics) that the project decided to only administer the Speaking Test to 10% of the population? What effect does this have on reliability of the results?

Woodlen replied that this was for the very realistic physical and logistical problems inherent in attempting to collect speaking samples from a wider group. The collection of samples of itself was not a very difficult thing but one envisioned platoons of people sitting around for a summer listening to evaluate those samples.

Woodlen wanted the group to bear in mind in forming judgments this morning is that the group is discussing the state of the art five years ago, not what has been developed since.

Valdman agreed that the tests or the scoring have not been modified in the last few years. He still wanted to know how a 10% sample of the population weaken the conclusions drawn.

Berger stated that a 10% randomly selected sample is fairly good. Secondly, the control introduced in the testing situation was rather rigorous. It was a one-to-one situation in a private room with no interference and administered by trained testers. This required about twenty minutes to one half hour per student; the procedure would become unmanageable for a large sampling.

Woodlen said the teachers' bias was completely ruled out because the project representatives selected the students to be tested at random.

Valdman believed that except for limitations due to the equipment of the school it is no more difficult to administer the Speaking Test than the Listening Comprehension Test.

Woodlen stated that the project testers did not depend on school equipment. Each of the testers had his own tape recorder. To insure uniformity and fidelity, all speaking tests were done on the same type of tape recorder.

Roberts asked if it was not true that the Speaking Test not only tests vocabulary control but has many different parts: mimicry, critical sounds, a global rating for intonation, picture question, picture description and picture sequence.

Valdman pointed out that there are a small number of sentences used to determine pronunciation. Three different scores are given for each sentence. There are only two major components to this test, one is pronunciation, the other is the ability of the student to produce utterances given a pictorial stimulus. Whether or not this is slanted toward vocabulary is a question open for discussion.

LUNCH DISCUSSIONS

The following points were reported to have been made during the lunch recess.

Carroll questioned the use of single analysis of variance of groups with correlated variables; the associated Tukey "A" for these analyses he believes, are irrelevant. /See Student Opinion sections of both reports for examples of this type of analysis/

Dayton, Hayes and Smith concurred that this might not have been the best procedure. Smith pointed out that it was done in the interest of economy since one analysis of the Student Opinion Scale data required approximately four hours of computer time. Many such analysis had to be made during the course of the study.

The computations were done simultaneously but no conclusions are based among both groups and administrations at the same time, only among groups on the same administration.

Roberts pointed out to his luncheon companions that the strategies, as defined and implemented, really were distinct. (See pp. 6, 7 and 8)

Smith mentioned that the real difference between the strategies was not so much one of English versus the foreign language but more properly one of cognition versus deduction - the "Traditional" strategy was predicated on student knowledge of structure with subsequent manipulation; "Audiolingual" was predicated on student mastery and manipulation with grammar being presented inductively. The question then, was not one of the amount of English so much as the ability of students to learn more when they recognize to some extent what they are doing.

AFTERNOON SESSION

Smith opened the afternoon discussion by citing Benjamin Harrison's remark on progress in education, "...a state of peaceful calm without friction is likely to mean either that nothing is going on or that what is going on is so far removed from the significant events of life that it doesn't matter."

Smith hoped that the discussion would center on the statistical treatment. It had been brought out that most discussants are in the applied areas. The research specialists will be able to tell us if the statistical treatment employed in the study answers many of the objections about controls.

The use of class means is a most meaningful unit. Campbell and Stanley make the point that if in large scale real-school research, significant differences do show up, it is more meaningful than in test tube situations. Most of the concerns voiced this morning are taken care of in the statistical treatment.

However, before discussing the statistics, the group returned to the definitions of the strategies. Smith thought that these were developed at the beginning of the project by the people planning the study (which included Woodlen, Berger and Roberts) along with the Pennsylvania State Foreign Language Specialists.

Woodlen stated that the panel of consultants devoted nearly a half day developing the criteria for the strategies. A tape recording and notes were made of the discussion. The definitions that were finally written and placed in the teachers manuals were the result of the discussions among the six or eight of us in Philadelphia that June day. In the concepts expressed, Woodlen was merely a vehicle, an amanuensis, in the strict meaning of the word. There was considerable emphasis on the need to be able to pull apart these various teaching strategies and establish polarities. It was very fuzzy in the first document and there was a very strong effort made to separate the various strategies in terms of pupil and teacher behavior.

Berger reviewed the responsibility of the consultants. They were presented with a funded document that could not be changed because of the contractural commitment to the U.S. Office. Their responsibility was to attempt to refine, define and to help implement the document. The original document was a plan, a blueprint, in a sense. The consultants were not asked to respond to the research design and statistical treatment.

Smith asked if the criteria and defintions developed at that time were regarded as precise statements of "Traditional" and "Audiolingual" teaching strategies.

Starr could not remember writing a list of general criteria.

Valdman remembered the discussions on criteria and pointed out that from the very beginning the consultants were concerned that it was very difficult to identify three strategies.

Berger stated that the consultants were not asked to respond to the final document.

Smith asked the group to turn to the statistical treatment of the main objectives: which was the most effective method? and did the language laboratory as used have any effect?

Dayton said that it was very difficult to briefly state what statistical work was done because there was a fantastic amount of analysis carried on. Some classes were at various points deleted from the final analysis. Classes which were not randomly assigned to groups were compared with those that were. They were deleted from the analysis of the first year data and stayed out through the analysis in later years.

Most of these factors which could be contaminating were one way or another eliminated. Classroom means were used throughout, but in many of the analysis data from individual subjects was used when he treated the design as nested. For example, the actual analysis as far as the comparison across columns of the table (systems) involved comparing class means or utilizing the means as a single score.

At the same time he looked at the performance of the students associated with the individual teachers. There were very large teacher differences. Indeed, in some cases the teacher differences are so large within a single treatment group as to completely overshadow comparisons across opposite dimensions. The difference between two teachers is, on an average, greater than the rate of difference between the treatments one tries to impose.

Covariance analysis was used throughout. All of the differences reported are post-measures adjusted in terms of initial level. A chance classroom in which all students are from bi-lingual homes becomes irrelevant. After adjusting for initial differences (which is presumably influenced by the bi-lingual upbringing) the contribution made to the criterion measure has been accounted for.

In such cases there is no danger to overall conclusions of a study of this type. There is probably a major difficulty in trying to apply such a study in real situations. Internally, considering the kind of randomization that took place, one can not criticize the results with respect to the group of participating schools, given the materials that they worked with.

Carroll asked about the publication of complete statistical analyses.

Smith pointed out that the reports only contained selections of pertinent analysis that related most closely to the primary objectives.

Dayton stated that he would hesitate to print the results on the teachers' nesting, since there are no valid error terms to compare results.

Carroll asked if an analysis of the total variance was made to obtain proportions of variance assignable to various effects.

Dayton said this was impossible due to the various sets of variables to analyze. Proportion of variance on a single variable is easy to find on the print-outs. Total sets of criterion variables, in the total proportion of variance accounted by rows, by column, or by predictors is also in the print-outs. In all cases computation began with an initial multiple variance analysis with up to twenty-five variables. The analysis probably could have been fabricated on a regression program.

Smith stated that this was being done for Levels III and IV.

Carroll remarked that sometimes it is more meaningful to do an analysis with a relatively small number of variables known to be significant.

Studies of this type do not give the treatments effects very much room to play around in, he believed.

Dayton remembered that in terms of the magnitude of typical pre-post-correlations, they were not high enough to be concerned. The typical values from .5 to .6, leaving sixty to seventy percent of the variance unaccounted for. The teacher proficiency measures did not seem to be particularly predictive of student performance and were omitted.

Lado stated that he was concerned about other things that a teacher contributes which are not necessarily measured by the MLA Proficiency Tests.

Dayton pointed out that such a refinement of the data were not available.

Lado still was concerned about the possible imbalance factor thrown in by native students, second generation in German and in French. The scores such students make at the beginning may not be very significant due to a store of dormant knowledge. The study did have randomization but Dr. Lado was not convinced that this accounted for all student and teacher factors. The results find a statistical difference between these groups but one still does not know what caused the difference.

Dayton said that Lado was raising the primary question of Type 1 Error--that results are due to chance. This is possible in any statistical study.

Higgins reminded Lado that randomization will insure that the probabilities of individual students with latent skills are uniform across treatments. The utilization of class means tends to nullify these effects if the randomization was by some means skewed.

Dayton added that such persons would also be removed by the analysis of covariance. It would be very odd if students whose previous exposure to the foreign language was not reflected in pre-measures, but during the year, he balloons. That one would have enough of these to influence a class mean and they happen to be concentrated more in one treatment than another is a confluence of chance factors which could happen, obviously. One has to find some probability which is very low to allow this possibility to influence thinking.

Woodlen remarked that the reason the Cooperative Test was used as a pre-measure was to endeavor to pick up youngsters who had some proficiency in the language. As a native Pennsylvanian, he was quite aware of this possibility, particularly in German, existing in the Allentown area and Lancaster County areas. Individual pupil scores of the classes in those regions were examined to see if there was any evidence of a preponderance of this type of youngster in the class. There did not seem to be anything in the data that suggested that there was a need to examine this more carefully. Individuals would become submerged since the study operated on a class mean basis.

Dayton clarified the reasoning in utilizing identical measures as both covariate and criteria. As long as this is done for all groups, presumably the sensitization and the learning that take place as a result of taking the test is equal. It is very unlikely that a student even remembers a single item from the first test for ten months anyway but this does not matter as long as it is done for all the groups.

Woodlen stated that this was exactly the logic followed when the study was designed.

Higgins asked Dayton if he were in a position to interpret the significant F-ratios obtained as being a function of treatment factors. The analysis apparently indicates that there are differences in certain criteria between the various groups which are beyond those attributed to chance. The obvious influence should be the treatment factor.

Dayton agreed that it is reasonably apparent that there is some difference between "Traditional" and "Functional Skills" strategies. Exact treatment and biases seem to be pretty accurately controlled. Whether or not one can reproduce that difference again, of course, is a crucial question in any research. Obtaining those same differences by applying these methods again and again is the real mark of success.

Smith and Woodlen pointed to the replication study which used the same treatments and teachers on a smaller scale with the same statistical analysis and the same results.

Dayton said that in broader terms, the study treats certain schools in the State of Pennsylvania. The ultimate test is whether this carries over into Maryland and elsewhere. He tended to agree with those who supported a more careful analysis of treatments.

Carroll remarked that the numbers are different but that the difficulty lies in their interpretation.

Dayton reiterated that differences were due to "something" that happened to "Traditional" classes versus to "something" that happened to "Functional Skills" classes. The specific mechanism should be the next concern. Researchers in this field should want to know what did make the difference. This question cannot be answered from treatments which, by necessity within a large scale project, cannot be controlled to the degree that you want.

One must quantify treatments. Comparisons involving qualitive treatments do not go beyond the first stage.

Berger asked for examples of quantification, vocabulary control, word counts both on treatment and also on the criteria. One can count hours of certain types of instruction, for example, X number of hours spent doing certain kinds of pattern drills.

Dayton agreed that might be a quantitative level but was thinking more of controls on quantifiable aspects of the treatment rather than trying to figure out what they were after the fact. How these strategies would fit on a physical scale so one could have not two levels but a hundred--although more realistically, four or five.

Carroll reasoned that this would no longer make a comparison between just those two rows of the cells diagram but really more with quantative variables. This certainly would be a very useful thing to do.

Berger felt the study should move in the direction of the FSI levels of skill from the standpoint of criteria measures. Instead of means and single scores, identification of a continuum of skills and describe it functionally. The criterion measures could also be plotted along some continuum.

Valdman pointed out that the problem with the FSI test is that it is not very reliable in the low part of the spectrum. It is reliable in the middle and less reliable in the lower and higher parts of the spectrum.

Carroll did not think that there has been an equation of the MLA MA Forms in French and German with the FSI ratings but he has made this equation for Spanish. Even your "Traditional" classes that are supposed to be doing best on the Listening and Reading Tests--if you were to use the Spanish norms with the same numbers (which is illegitimate)--they would still be only an S1 level. Even after three years, none of the foreign language students in this study do very well, you might say.

Valdman was amazed that even after three years students did not reach the S2 level.

Carroll said the mean is at about the S1 level if the construction of the Spanish tests is more or less comparable to that of the French and German.

Smith said that earlier he detected a feeling among several discussants that the three treatments were, in effect, the same. Why then the differences at the end? This does not seem logical.

Carroll now raised the question of the difference between statistical significance and practical significance. He saw very little practical significance between the TLM and the other treatments because there is only about three points on the raw score scale in typical data. This is very small with reference to the total range. What struck me about the whole thing was the fact that even after the three years now the students were not doing very well.

Smith stated that he had visited the classrooms and observed the students. From his experience as a State Foreign Language Supervisor, the students in the study seemed rather typical students. This is why the study questioned the published MLA test norms. They are also questionable because they have never been updated despite the availability of much data. Norms still reflect small early samplings.

Carroll mentioned he had norms on the MLA Proficiency Tests.

Smith asked if they have ever been redone from original norms based only on NDEA Institute scores.

Carroll said "No."

Smith remarked that project teachers compare favorably with national NDEA Institute percentiles, the fiftieth to sixtieth, but the students in the whole study did not compare well with the ETS norms. Were project classes not typical of Pennsylvania?

Eisenstadt observed that the Booklet of Norms /MLA Cooperative Classroom/ indicated that there were a fairly sizeable number of private schools and academic high schools included in the Pennsylvania contribution to the norming population. In addition, in 1963 there were not many truly "Audiolingual" classes. ETS had to take the teacher's word when they submitted test scores for class performance that these children were really taught audiolingually all the way down the line. He questioned how many people had attended institutes and had pre-service training in "Audiolingual" teaching before 1963.

Glaude asked if project schools were representative of Pennsylvania--that they represented normal students?

Eisenstadt reitereated that the project sample included urban schools, suburban schools, and rural schools. They were not concentrated in any one area, but were spread out over all types of population and should give a fair cross-section of the schools in the state. There was a geographical concentration in the eastern part of the state but not in the city nor solely in the suburbs.

Carroll observed that ETS in making norms, would normally have a policy of trying to include both public, private and parochial schools. In several of his studies quite wide differences between public and private schools in foreign language were found. Project norms are more appropriate for public schools whereas the ETS published norms are appropriate for mix of public and private schools.

Woodlen commented that the norming population on the MLA Cooperative Classroom Tests was extraordinarily small for a test of this scope. Some of the samples were below one hundred on some parts.

Carroll agreed.

Higgins asked what proportion of project teachers had been to NDEA Institutes.

Smith replied that forty percent of project teachers had been NDEA participants while the state average was twenty percent.

Higgins asked what this meant. Were they from atypical districts? More progressive? Closer to institute sites? What?

Smith commented that there are many factors involved in the selection of NDEA participants. It indicated on the whole that they were teachers who were interested in improving themselves.

Starr added that theoretically they should be more knowledgeable about the so called "Audiolingual Method." But before extrapolating anything at all, one would want to know which institutes they went to and what level of institutes they were. If one jumps suddenly from that "40% NDEA Institutes, therefore..." and there is no "therefore" unless one knows other things, too--which ones they went to, how they scored there, what level they were when they went and when they came back.

Smith mentioned the point needed to be made that some of the "Traditional" teachers had been to NDEA institutes and that this factor was scattered across the strategies.

Woodlen thought it relevant to realize that the nature of the study deliberately searched out the schools and staffs that had the language laboratory facilities. Therefore, there was a kind of implicit selection in that the teachers in those schools may have responded to pressures to attend NDEA institutes.

Carroll asked why the study excluded teachers who had been to the foreign country over the last two years.

/Reference to Teacher Control, Point 2. Report 5-0683, p. 27. "...teachers who had recently spent considerable time abroad (two or more years) in the country where their foreign language is spoken were excluded."/

Berger suggested that these teachers were avoided in the experiment because the teacher factor would dominate the treatment. It would have been a contaminating factor. There were no native speakers among the French teachers and only a few in German. Native speakers were excluded from the teacher population unless they had been residents of the United States for many years.

Smith reviewed the decision to use, as an experimental variable, two one-half hour language laboratory periods per week, about twenty half hour periods two days a week. Some record, some do not. The results of the study indicate that there is no significant difference between those who go to the language lab and those who do not. There is also no significant difference between those who record and those who do not record. The schools in which the study was done have not yet modified their use of the laboratory.

Lado objected that the measuring instruments weren't fine enough to find out the differences. /MLA Cooperative Classroom Tests./

Smith pointed out that the researchers tried to find a suitable test in the widely accepted MLA tests. The project also commissioned Rebecca Valette to write special tests which look promising but need further refinement.

Lado believed that the study then did not permit the conclusions drawn.

Smith remarked that the conclusions reflect only the instruments used--on the tests that were used no significant differences were found. The text of the report reads-- "...the language laboratory, as employed in the experiment, had no discernable effect on these measures."

Lado hoped that out of the discussion at least one person said the measuring instruments were not adequate, therefore, there just was not a way to find out differences. If the MLA tests are regarded as universally accepted by the profession, the conclusions may be taken as being final by readers.

Valdman added that the measuring instruments might be biased. For example, the Listening Comprehension Test may not measure true listening comprehension, the speaking test may not measure true speaking ability.

Carroll supported this as a very important point.

Dayton asked Lado if really relevant instruments were available, what differences might have been found.

Lado commented that perhaps the use of the language laboratory would have shown a difference.

Dayton reminded Lado that perhaps the use of the laboratory would definitely be much worse since when failure to find a significant difference indicates a state of no information.

Dayton went on to state that it is very, very unlikely that given any three treatments (of any kind) that in the long run there is no difference in any variable one wants to name.

The problem is one of ordering the three [TR, AA, AR.] The fact of the matter is one cannot order them on the basis of these variables [TR, AA, AR.] There is no way of ordering it since no significant differences exist.

A choice must be made on some other basis then the kind of outcomes measured by these variables. This might be in cost, in convenience, it might be almost anything.* The state of the information is such that one cannot conclude--because there is not a significant difference--that what one did not find was a favorable difference *for* the lab.

*N.B. See Specific Objective 2, Report 5-0683, p. 10.

Carroll mentioned a study that he did at least ten years ago which came out very much in line with the [Pennsylvania] study. Students who had quite a bit of language laboratory experience were no better, in fact were poorer, than students who were with a teacher who used a lot of language and phonograph records. On reading tests and tests that had vocabulary the students without the language lab were better. His interpretation of it was that the student that was without the language lab had much more chance to read to be exposed to language both orally and written form. They were able to acquire a better vocabulary and better mastery. The outcome of the Pennsylvania study is not at all surprising.

But he thought that the group should guard against saying that the language laboratory was no good. It may have been a good supplement or a good substitute for teacher deficiencies and the fact that you find no difference between these two efforts--between language laboratories and non-language laboratories--should not be taken as a condemnation of language laboratories necessarily. That is the impression that is likely to be purveyed by some of the publicity.

Smith agreed that people are learning of the reports and saying that language laboratories are no good. The reports never make this statement.

Valdman asked if the record versus non-record comparison took into consideration the type of operations and exercises that students undertook in the language lab--the type of materials to which they were exposed? Perhaps proper use was not made of the recording facility.

Smith stated that students who started out in a recording treatment continued to record for two years. There was never any back and forth between record and non-record. These people always recorded the laboratory exercise. Students who were in the audio-record treatment never recorded except when they were individually interviewed for the Speaking Test outside the lab.

Valdman would like to have some quantifiable factual data. How much did they really record? Maybe that up to a certain level something was significant.

Note: Text of the reports is not precise on laboratory usage. Refer to "Teachers' Guide Materials" Report 5-0683, pp. B-7 and B-14.

Smith reviewed the fact that students were assigned to record one-half of the laboratory /25 minutes/ period and then to play it back one-half of period. In theory, students listen to themselves and correct their own errors.

The great flaw of the Keating report was the materials used. Materials used in the Pennsylvania project were those made available through Holt, Rinehart and Winston and Harcourt, Brace and World--those that the average teacher has. Teachers were not allowed to be creative. It assumes that these audio programs were created by "experts"--the best that money could buy.

Esseff asked if there were more strict controls on the physical operation of the lab than is evidenced in the report. Some were inoperative, but he was thinking in terms of physical operation. Were any of the labs dialed? Were there any attempted comparison between size of the lab or the condition or the fidelity? Or a whole host of other variables--climate controls, the location of the lab, the number of student positions? In general the study indicates a type of laboratory that supposedly has an audio response but it is not further defined.

Smith believed readers have too narrow a view of the study. This is a curriculum assessment much more than an experimental study--this is what schools are really like. The language laboratory breaks down tonight and is not always fixed by tomorrow morning.

Eisenstadt observed that the project did gather a great deal of valuable information about the equipment in the laboratories, the age of the lab and teacher training and maintenance. All this is on file here and in Harrisburg but was too voluminous to include in reports.

Kilchenmann asked if any labs were open during study periods where students could study on their own, on their own time, at their own speed?

Woodlen replied that there were not. This could have destroyed the prescribed treatment.

Lado again defended the "test-tube" experiment as opposed to the "mass experiment." If, in a "test-tube" experiment one finds out that the lab used in a certain fashion does produce something--then later you find out in a massive experiment that it is not producing--then one knows why it is not producing. If the large scale study is done first, one then cannot isolate contributing factors.

Valdman asked if the students were trained to use the recording possibility. He has found this to make quite a difference.

Lado added that there is another factor that is impossible to measure--a very solemn one. He was instrumental in starting a program in Spain by which Spanish universities developed English departments for the teaching of English. In Spain this did not have a tradition. He had suggested that whenever the Spanish started an English department they install a language laboratory. Lado admitted that he had always been very skeptical about the language lab--yet in Spain it had a very specific purpose. The moment there is a language laboratory in the university that has an English department, the reason for learning to speak the language does not have to be defended.

Lado felt that language laboratories in the American movement have contributed a great deal to establishing the desirability and even feasibility of teaching students to speak, whereas Coleman had concluded it was impossible and threw it out.

Zimmerman asked what was done other than just play back? Were there criticisms from the teacher? Recording for the sake of recording does nothing.

Smith agreed but reminded the group that there were at the time those who said that students are capable of self-evaluation during playback.

Note: The teacher did monitor and correct students. Report 5-0683, pp. B-17, B-18.

Kent inquired if recording would not improve the speaking ability and sound production.

Smith stated that it did not do so meaningfully in the study. There was no difference. In the critical sounds area the students who did not go to use the laboratory at all did better.

Hilaire observed that there is no way to measure fluency. The tests do not measure how long it took to answer. Supervisors see the students that are in both "Traditional" classes and "Audiolingual" classes. The big difference is in fluency.

The reasons why students are better in "Traditional" classes when given tests they usually are given are (1) because the tests are not necessarily testing what is emphasized in "Audiolingual" classes; and (2) that the students still in third year "Traditional" classes are very good--those in third year "Audiolingual" classes are average. Everybody in a third year "Traditional" class would also be in a third year "Audiolingual" class--but there are a lot of other kids in the third year "Audiolingual" class who would probably not have made it in a "Traditional" sequence.

Halaire believed that the student questionnaires are questionable. People lie on questionnaires. Students cannot compare strategies.

Smith said that students were not asked for a comparison but for suggestions on how to make their own course better. Interviews were trained by a guidance specialist to note things that were characteristic of "Audiolingual" or things that were characteristic of a "Traditional" approach.

Carroll observed that in looking over these reports he did not see any studies where investigating possible interaction between aptitude and treatment, one of the hottest subjects in educational research. The study must have data of this type. Chastain at Purdue did such a study. He compared two methods, the "cognitive code-learning" method and the "habit formation" (Audiolingual) method. He found that there was an interaction between aptitude and treatment. The students in the lower ranges of the aptitude measure were better under the "Audiolingual" method whereas the kids in the upper range of the aptitude method were better under the more "Traditional" cognitive method.

Dayton asked whether Chastain had used post hoc blocking, forming the blocks, the hi-low, after the end of the experiment.

Carroll stated this was the case.

Dayton believed this would make it impossible to make that comparison. One of the assumptions of lot comparisons of students is that one randomly assigns across treatments among blocks.

Smith pointed out that the project did investigate the relationship of intelligence and aptitude scores to strategy. /Report 5-0683, pp. 80-81./

Hilaire asked about failure rates in different strategies.

Smith reported that nowhere did the project use as data teachers' subjective judgements or grades. It was made clear this morning that students were excluded for whom the study did not have experimental data. He could not keep track of every student. Most students at the end of Level II felt that they had completed the two years necessary for college entrance. Numbers of students continuing in the experimental population are not relevant because this does not allow for moves, transfers, and many other factors the staff could not count.

Smith asked that the group discuss the findings of the study.

Esseff asked why did Smith make Recommendation 7 on page 114 [Report 7-0133].

Note: The recommendation questioned reads: "That future educational planning envision the language laboratory in terms of individualized practice in addition to regular classroom instruction rather than as a type of classroom activity."

Smith replied that Recommendation 7 was written because the lock-step laboratory does not seem to work. The next logical step is to investigate other ways to make the language laboratory more effective.

Smith observed that every single conclusion refers to a specific table in the data analysis. There is not anything interpretative about the conclusions. They are factual. For every conclusion one can point to data.

Lado asked why then had Smith indicated "some significance" in the production of key foreign language sounds. What is meant by "some" significance? It is either significant or not significant, and it is significant at one level or significant at another.

Smith conceded that was a good point. It should not have said that there was [Conclusions, Objective 1b, fourth line] "some significance" in the production of foreign language sounds on the unvalidated Valette test. [Table 24, Report 7-0133, p. 66]

Valdman pointed out that in the conclusions for Objective 1 it is stated there is no significant difference in listening and reading skills. But in speaking and writing there is no significant difference as established by specific tests. One then could infer that presumably there is some reason to state that the differences in listening and writing were reached on a basis other than instruments.

Smith admitted that the omission of the mention of specific instruments was to avoid repetitious statements.

Valdman believed it would be important to point out that there is or is not any difference in a skill as measured or as established by a given instrument.

Carroll added it would be better to have a general statement at the beginning saying that all these conclusions should be qualified in terms of the particular instrument used.

Valdman said that conclusions are as valid as the tests are valid.

Esseff referred to Objective 2 on page 110 (7-0133).

Note: The text of Report 7-0133 reads:

> To determine which of three language laboratory systems is best suited, economically and instructionally, to the development of pronunciation and structural accuracy.
>
> In Level I, Project 5-0683, no significant differences existed in foreign language skills classes using (1) a tape recorder in the classroom and those receiving additional practice twice weekly in either (2) an audio-active or (3) an audio-record language laboratory. At the end of Level II, significant differences between these three groups failed to emerge. The language laboratory had no discernable effect on listening or speaking but laboratory time may have influenced reading skills.

Esseff believed that this conclusion as it stands says more than it should say. The unrefined labs as used in this experiment were gross. The conclusion and the reports do not coincide. They should be more qualified.

Smith stated there was no significant difference among tape recording, audio-active and audio-record systems. Esseff is suggesting that one should have somehow checked on every single different kind of laboratory.

Esseff reiterated that it is grossly stated, grossly in a gross experiment. Defined in a sense but grossly evaluated. The reports refer to three language laboratories systems when in reality there may have been fifty language laboratory systems. It implies more here than the study warrants.

Smith restated that they were defined as basically different "Systems," as discussed in works by Hocking, Hayes, Hutchinson, Stack, etc. Within each there are particular variations and arrangements.

Esseff stated that he reviewed language laboratories at the rate of one hundred a year. There are so many variations in the language laboratory system in higher education alone that he felt very uneasy accepting that conclusion /Report 7-0133, No. 2, p. 110/ without some qualification on variations.

Smith reiterated that the study assumed there were distinct "Systems."

Roberts added that qualifying statements on individual language laboratory installations would only be valid if one were making comparisons among individual language laboratories. The project made a comparison between a particular use of all these language laboratories taken together as a system against non-use. It is a gross comparison.

Roberts believed Esseff was referring to comparison as between one language lab with its particular set of conditions against another kind of language lab with its particular set of conditions. That was not undertaken by the project.

Esseff again stated that the study did not define any of those conditions. It impliesthat these conditions had no effect upon the results. They must have had some effect.

Higgins believed the researchers defined terms very well. The study is never going to be "idiot" proof.

Lado stated that as he read the conclusions they sounded absolute, far beyond the nature of the information that went into them. Editorially, they need to have several cautions: that questions have been raised about the adequacy of the tests, that questions have been raised about the adequacy of the differences in the teaching strategies used: questions have been raised on the controls. These reports can do a lot of damage.

Berger pointed out that the transcript of this discussion will be made available. The decision has been to attach to all releases of this report henceforth, a copy of these reactions. People who read the report will immediately read the responses.

Lado believed people are going to read these two or three pages of conclusions and a lot fewer are going to get down to the comments that each one of us made. He thought that the style of these conclusions is too absolute and is not justified.

Valdman questioned Objective 5 stating he thought the researchers said here much more than was actually intended. It refers to levels of foreign language mastery that are obtainable from the secondary school language program, yet the study did not really exhaust all the various possibilities; for example, flexible scheduling, programmed instruction, audio visuals, etc. Anyone who reads this would probably infer that these results are to be interpreted as what you really can teach. It would be easy to modify this editorially so that people do not read too much into it. This could be very dangerous if one believes this is about all students can learn in high school.

Smith stated that Objective 7 is the one that bothered him the most. There was no discernable relationship, even after three years, between teacher scores on the MLA Proficiency Tests and class achievement.

Starr suggested that care should be used in mislabeling this test battery the MLA "Teacher Proficiency Tests."* This is in the area of implication. In justice to the developers of these tests, it can not be said too often that no claim was ever made that they were measuring teacher proficiency in the sense of the effect of the teaching. They were never intended to be anything else but standardized measurements of what they claimed to be--the four skills and three content areas.

*Note: The tests are entitled the Modern Language Association Foreign Language Proficiency Tests for Teachers and Advanced Students.

3

It was probably an unfortunate title to give these tests. They should probably have been the MLA Advanced Tests to distinguish them from the Classroom Cooperative Tests.

Smith stated that for economy, the reports took the liberty of reducing Foreign Language Proficiency Tests for Teachers and Advanced Students into Teacher Proficiency Tests.

Starr observed that the recommendation that the Proficiency Tests not be used as part of the certifying process seems to say that the researchers do not really care about measuring the skills of teachers because one cannot measure the effect of these skills on students.

Smith believed that research to demonstrate relevancy should be done before the imposition of criteria for teacher certification.

Higgins suggested Starr brought out a major concern--what the students learned. The MLA Proficiency Tests instrument was not indicative of the extent of a teacher's ability to increase student performance, it does not make sense regardless of level. The statistical effects may be somewhat reduced because the study restricted the range of this instrument by state law, the bottom extremity was cut off. This may suppress some correlation. There was logical reasoning behind the recommendation.

Hilaire asked if there were any really minimal proficient teachers in the experimental population.

Smith answered that six of the project teachers could not have been certified if the state requirements for minimal scores on the MLA Tests had been retroactive.

A preliminary study of the correlations between teacher proficiency and class achievement indicate the possibility of an inverse or curvilinear relationship.

Lado asked if Smith really believed that knowledge of the target language is irrelevant to a good teacher of foreign language on the basis of this data?

Smith denied not being concerned with teacher skills but that he was questioning the instruments and their application.

Starr said what was really discovered in the basis of this research is that there is no predictability between the Proficiency Test scores and the success of students. That is what should be stated.

Smith commented this is what the report did say: that there existed few significant correlations.

Roberts reminded the group of the old adage, "He certainly knows his subject but can't teach."

Smith believed that the mandatory imposition of a criterion of teacher licensure is, in effect, telling teachers that the criterion relates to ability to teach--and it does not.

Lado maintained before one can say that, one would first have to have an experimental situation using a real representative group of trained and untrained teachers--teachers who knew the language and who did not know the language--and then correlate the two. Then if one came out with this lack of correlation Lado would accept the facts. The Pennsylvania study had a highly selected group of teachers. Elementary statistics teaches that if one has a highly narrow band and one cuts off the bottom, the correlations are going to be shot.

Smith reminded Lado that Higgins had just brought out this point--that the study cut off some people by state established scores. Actually it did not. Six of the eighty-nine teachers in this analysis would have failed the state tests had they been retroactive. The study did contain a wide range of teacher proficiency.

Higgins reminded Lado that he was talking about a population of teachers not a population of men on the street. In order to make Lado's suggested analysis meaningful, it would have to include as prospective teachers both teacher candidates from bacculaureate programs and totally untrained teachers.

Lado asked if the six poorly scoring teachers were chosen deliberately.

Smith stated that they were included in the sample. It was not known that they could not have been certified until we checked their scores two years later. They were accepted as being qualified to teach under existing state certification requirements.

Valdman believed that many questions that have been raised on the conclusions may be due to the way they are organized. It might have been helpful if each Conclusion and each Objective had been followed by discussion.

Smith said that the original manuscript was written as Valdman suggested but was re-done to comply with the USOE format: discussion and conclusions separated.

Valdman commented that the USOE format does not prohibit a report in which one tries to account for results. In fact one suggests additional research or the preparation of additional instruments. The report does not state that it would be useful to try to develop some finer instruments in various skills.* People need to know that, despite what many people think, the MLA Tests at all three levels could be improved upon and are not to be taken as absolute measures of proficiency for various skills.

*Note: An attempt was made by the project to develop finer instruments. Ref. Report 5-0683, pp. 39-40 and Report 7-0133, p. 37.

Lado reiterated that one could, for example, come to the conclusion (1) it is not important at all whether you have trained teachers or not--therefore henceforth no more training of teachers; and (2) language laboratories are no good--out go the labs. Pretty soon what do you have. Lado wanted to go on record as stating that the data does not warrant the conclusions.

Smith pointed out that he has been critized by a competent educational statistician for being too conservative on this. This professional stated there are enough negative correlations to show that these tests predict inversely--the better teacher scores on this, the worse the teacher. The reports avoided stating this.

Yanis reminded Smith that he was the research specialist alluded to and that he did not support the recommendation that the tests be dropped. This itself is not warranted by the results of the study. The results are a suggestion that teachers who do not achieve added proficiency in the use of the language are just as qualified to teach it as those who do.

Smith believed that the interpretation must depend on the definition of "adequacy," which levels have not yet been defined. The state scores were picked arbitrarily.

Kilchenmann believed that the study did not cut off the bottom but the top for "Audiolingual," i.e., study abroad, and this is significant, much more so for "Audiolingual" than for "Conventional."

Newton stated that despite USOE restrictions on the discussion, someone in this group owes it to the profession to write an article stating the limitations and restrictions under which the project labored. For six years under under Dr. Boehm, Pennsylvania went forward tremendously in language instruction. This study is going to set us back to the pre-World War II days. Despite the scores students have shown on these tests, at the college level they are coming far better than ever before. Teachers are overwhelmed with their knowledge of the intermediate courses and can place students in advanced courses when they demonstrate proficiency.

Note: Miss Newton was invited to describe the limitations she alluded to for publication in the SUPPLEMENTARY REPORT. They had not been received at publication date.

Hilaire and Smith both commented that longer sequences were a contributing factor to this observed improvement.

Woodlen reminded Newton that one cannot assume that these improved students are necessarily graduates of "Audiolingual" curricula.

Newton agreed that some are and some are not. Some are--but no matter which teaching strategy they learned in, language per se has improved because of all of this ferment. This project will set us back.

Valdman again stated that one reason that the study is disquieting is that perhaps it interprets "Traditional" literally. In fact, this is not a "grammar translation" approach but a "traditional-ecletic" approach with many improvements. However, most readers of this report will interpret "Traditional" in exactly that way. They will use this as justification for the practices of the last thirty years, saying "I knew these people were all wrong. As Chomsky has said, structural linguists and certain types of psychologists have sold language teachers a bale of goods, and there is nothing as good as old traditional."

Valdman continued to say the reasons that he and his colleagues, as members of the consultant board, would like to see a stylistic revision of the report is that the report is widely distributed and widely known. People may not read articles but they will read the report. Some of these caveats, interpretations, explanations, and restrictions should be included in the report. This is certainly compatable with USOE regulations.

Roberts emphasized that there has been a lot of talk about the definitions of the various strategies or methods. The definition of the "Traditional" mode was a definition of a good "Traditional" approach. In the last five years there has been a movement toward reassessment of the "New Key." This was no where more apparent than in the October, 1968, issue of Foreign Language Annals. Practically the whole issue was devoted to the idea of reassessment: now is the time to take stock, where do we go from here? and the profession may have gone overboard with the "Audiolingual" approach. If this project does nothing else, at least it contributes to that attitude.

Lado believes the Pennsylvania study contributes to throwing out all new ideas. Starr agreed.

Zimmerman believed with modifications the reports are very valuable documents--but leaving them as they are now, it says throw out the language laboratory, throw out this, throw out that.*

*Where? No reference. P.D.S.

It points out that the profession should create better tests, use the laboratory more effectively and improve instruction.

Starr warned that as it is now, ninety-five percent of the people that read it, scan it or refer to it in their articles are going to misinterpret the whole thing.

Smith stated that a certain amount of misinterpretation cannot be prevented by readers with particular biases. No matter what you tell people they will interpret it their own way.

Berger asked that a transcript of this record be sent to those present. They will in turn respond to it. If this is consistent with the actual experiment, the researchers will add the caveats because this is certainly editorial work.

Note: Of the twenty-four persons quoted in this document, 5 responded with minor corrections.

Valdman did not think this would help. He believed that when one undertakes research in this area, which is very sensitive, the people who report the research owe it to the profession as a whole to qualify their conclusions.

The things that have been said today, the modifications, caveats and so forth, are an integral part of the research. These should be part of a report. Simply sending the transcript of this discussion to other persons involved is not going to help very much.

Smith hastened to state that the discussion will be published in a third report and sent automatically to everybody who has ever gotten one of the first reports.

Berger believed the concensus of the conference participants is that the authors re-edit the conclusion section carefully in one of the following manners: (1) an introductory statement consisting of the following types of paragraphs: The results of this are limited, based on the instruments used, based on the labs as we found them during the years 1964-1965, etc. And (2) added recommendations that someone investigate the effectiveness of the language laboratory when used in "an ideal setting" with refined instruments; (3) the profession develop further investigations to find out whether or not instead of two times, five times a week would help.

Agreement by Lado, Starr and Valdman.

Baranyi asked since the research and the data are available, would it not be improper to include things that one would like to see come out of the report. For example, the suggestions on the language lab--what are the various differences?--the report did not contain specifications to test for those. It would be wrong, he believed, to allude to these in some of the conclusions since they were accounted for in the research.

One can only edit what has been done, not bring in other effects that have been discussed and learned about since the beginning of the project.

Berger agreed that the language laboratories portion of the study was grossly done. The researchers know of laboratories where the contract for repair was not in force for half a year. The study revealed many poor maintenance situations. These things should be investigated and should be recommended in the recommendation section.

Note: See Recommendation 6, Report 5-0683, p. 113. "A more careful and sound policy of language laboratory administration and maintenance be immediately initiated by responsible school authorities."

Roberts asked if what was being suggested is that in each one of these conclusions--besides stating what was found--it should also state specifically what was _not_ found to head off any unwarranted conclusion on the part of the reader.

Valdman clarified that he meant in the matter of a "Discussion" to say that "the reasons such-and-such was found or not found may be due to the following factors: (1) this is very complicated; (2) it is very difficult to conclude (3) one should be very careful in how one interprets from this."

Smith felt disturbed since he felt that somewhere the reports _had_ said all of these things. There is a great deal contained in the reports and maybe one cannot expect everybody to read every part of every research report. Maybe one reads selectively.

Valdman believed if one reads these reports carefully, one cannot but come away from the reading with the impression or the doubt that it was very difficult to isolate three strategies; that, perhaps, there were flaws in the control and, maybe, what one really had is one strategy which varies back and forth--the differences due to chance and factors which have not been isolated.

Esseff commented that he had reread the _Keating Report_ before the conference for curiosity. The last paragraph in that report (page 39) qualifies the results. No one seems ever to have read that paragraph. The Pennsylvania study is very valuable if there are ways of getting that value out without clouding the issues.

For example, on the language laboratory controls, it is certainly permissable to state that certain things were not investigated. People are more sophisticated than ever before in reading these reports and will accept things if parameters are included. There is no basis for making any generalizations without knowing what was done or not done in regard to a particular treatment.

Smith stated that he receives numbers of comments which clearly show that people have not thoroughly read the report.

Higgins observed that as a research specialist he had read the reports for a scientific viewpoint. He did not forsee the extent of the subjective reactions expressed by foreign language educators.

Smith closed the meeting by stating that a complete transcript of the meeting would be sent to all participants and that a condensation of the remarks would be contained in the SUPPLEMENTARY REPORT of USOE Project 7-0133.*

Several discussants promised to submit further comments and suggest implications of the study by mail.**

*Set to participants May, 1969-P.D.S.

**None received by 8/12/69 - P.D.S.

Evaluation of PDE/FL Research Projects

by

Albert Valdman

Clearly the most vulnerable aspect of the research is the establishment of the three teaching strategies and control of adherence to assigned strategies on the part of participating teachers.

1. Defining Criteria of Strategies

In the final report of Project No. 5-0683 (Jan. 1968), there does not appear to be any difference in the defining criteria of the FSM and the FSM + Grammar strategies. In addition, the categories into which the criteria have been organized are not always comparable. For instance, "vocabulary" appears only in the list of TLM criteria and "use of target and native language in classroom" and "sequence of learning" appear only in FSM and FSG lists of criteria. But most importantly, the criteria are stated in sometimes vague and imprecise terms, and this makes evaluation of adherence to the particular strategy of the part of participating teachers difficult indeed.

a) Vocabulary If TLM is characterized as presenting primarily learned vocabulary in terms of word-for-word equivalents rather than contextual equivalents, then one would assume, on one hand, that FSM and FSG present little "academic and literary" lexicon, and, on the other hand, that TLM presents little everyday "functional" lexical items. But an examination of three French texts that represent the three strategies (Dale and Dale - TLM; A-LM - FSG; Holt Series - FSM) shows that all three are constructed around dialogs and contain primarily everyday lexicon. A more useful criterion might have been size of vocabulary in the various textbooks used.

b) Grammar The only variable that distinguishes the FS strategies from TLM is the role of grammar in FL learning. In TLM, understanding of grammar rules is considered essential to the control of the behavior characterized by these rules, whereas, in FSM and FSG, grammar rules are considered "incidental". However, the latter criterion is contradicted, for FSG, by the "Rationale" which appears to state that intellectual understanding speeds up the acquisition of language habits.

The other critieria listed for "Grammar" are absolutely meaningless and they do not help differentiate the strategies. In all texts utilized, grammar is prescriptive inasmuch as the language used is characteristic of a standard form. The style of presentation of grammar differs only in that FSG is more explicit about the phonic manifestation of grammatical features. All three approaches present syntax -- by far the most important part of grammar -- in the same way. With regard to paradigms and lists, the strategies differ only in that in FSG, students manipulate forms in phrase- or sentence-long utterances.

c) Testing The reports do not make clear on what basis grades were awarded. Were experimental measures used for that purpose? It would be helpful to report on the nature of the tests FSM and FSG teachers used to evaluate listening comprehension and speaking ability and to what degree these tests contributed to final grades. One would challenge the assertion that dictation tests are essentially a feature of TLM. On the contrary, they constitute a broad test of listening comprehension and they may be used to test phonemic discrimination.

d) Use of NL and TL languages in classroom Clearly in all three strategies both teachers and pupils used the English and the TL. What is significant is the proportion of TL to NL use by teachers and pupils and the purpose for which each of the two languages was used.

e) Reading It is doubtful that in FS strategies the pupils never were asked to read material which they did not control orally.

f) Writing Surely in all strategies the relationship between sound and letter was pointed out to the learner. Indeed, one suspects that if such activities as dictation and reading aloud material not fully under the active control of the learner were considered features of TLM, then learners taught by that strategy would be more proficient in converting letters into sound in a language like French whose orthography does not provide a one-to-one relationship between sound and letter.

g) Sequence of learning As it is stated in the reports (e.g., p. 21 Jan., 1968), it is doubtful that there was an appreciable difference in the order of presentation of skills in the three strategies for any single structural feature. In both TLM and FSM/FSG, the passive skills (listening comprehension and reading) precede the active skills (speaking and writing). The only difference, then, is that in TLM, pupils are expected to learn to recognize visually grammatical features and vocabulary items they do not yet control audiolingually. But since FSM/FSG pupils were not deprived of access to the written representation of grammatical features, one must assume that actually they did not always progress according to the hearing - speaking- reading - writing sequence. In fact, if the audiolingual proposal for the sequence of skills is interpreted correctly, pupils should only manipulate orally material they understand perfectly, and there should be a time-lag between the auditory introduction and the oral manipulation of material. It is well known that this is far from being the case in many FSM and FSG classrooms.

2. Rating Scales

The rating scales do not always reflect the criteria which are assumed to distinguish the learning strategies from each other, and one suspects that the evaluation of adherence to assigned strategy which depended on their use did not effectively rule out contamination of this key variable. For example, with regard to vocabulary the rating scales only tell us that in both TLM and FSM/FSG there is some form of vocabulary drill and that in the latter strategies words are presented in context. But it does not tell us how vocabulary is presented in TLM. It is quite doubtful that in that strategy words are only taught in lists. We can only conclude that the manner in which vocabulary is presented is not a significant criterion in distinguishing between strategies.

Some of the categories in the rating scales appear to be meaningless or difficult to interpret, if not downright puzzling. Thus the TLM scale refers to "pronunciation" on the part of teacher and students whereas the FSM/FSG scales refer to "speaking the TL" on the part of teacher and students. One would infer, no doubt wrongly, that in TLM more than 3-5 minutes per day is devoted to pronunciation drill.

Perhaps it would have been more useful for the evaluator to use a single scale applicable to all three strategies. The scale would consist of a set of criteria for which scalar evaluative judgments (qualitative or quantitative) would be made, for example:

			Rating		
	(HIGH)				(LOW)
	1	2	3	4	5
(1) vocabulary list drill					
(2) vocabulary presented in context					
(3) use of FL by teacher					
(4) use of FL by students					
(5) use of NL by teacher					
(6) use of NL by students					
(7) pronunciation drill					
(8) formal grammatical explanation					

Whether the teacher adhered to the assigned strategy would be determined by the overall score on the scale. For instance, one would expect TLM teachers to score low on criteria (3), (4), and (7) but high on criteria (5), (6), and (8); for FSM teachers the scores on these items would be reversed.

To put it differently, the reports do not provide negative information: to what extent did teachers assigned to a given strategy engage in activities characteristic of some other strategy? Another potentially contaminating factor is the attitude of participating teachers toward the three strategies and their assignment of textbooks to the three strategies. It would be of some significance, for example, if teachers assigned to FSM held views toward FL learning characteristic of TLM, or if a teacher using a textbook defined as essentially TLM actually considered it suitable for FSM.

Some Conclusions to be Drawn from the Pennsylvania Study*

Rebecca M. Valette
Boston College

The publication of the Pennsylvania project report raises a variety of questions. The results of the first part of this project which point to conclusions, other than those many teachers had expected, means that the project will be analyzed with-a-fine-tooth-comb to uncover flaws in the design and weaknesses in the execution of the project. But despite possible imperfections in the research, we cannot ignore the findings of the study. We must admit that the teachers of the Commonwealth of Pennsylvania are probably no better and no worse equipped to teach foreign languages according to a method assigned them than teachers in other states. The language laboratories in Pennsylvania are used much in the same way that they are used in other states. Students throughout the country are given the MLA Cooperative Tests. What then are some of the questions we must look into?

1. Is the "Traditional" method superior to the "Audiolingual" method? The question as it is worded here is much too broad. The conclusion of the report is that first-year students of French and German taught by a "Traditional" method (as defined by the consultants) performed better than first-year students taught by "Audiolingual" methods on a specific set of tests: namely, the old Cooperative Tests, and the new MLA Cooperative Reading Test and the Critical Sounds section of the MLA Cooperative Speaking Test. It was to be expected that the "Traditional" students would do better on the "Traditional" Cooperative Test of grammar, vocabulary and reading. But how can we interpret their performance on the new MLA Cooperative Tests? The key to the reading test is vocabulary load. If we look at three of the texts used in the French classes involved in the study (i.e., the A-LM materials, the Holt materials and the Dale & Dale text) we find that each unit contains roughly an average of 50 new lexical items. The project report states that A-LM classes finished about 10.5 units; Holt classes finished 13 units and the "Traditional" classes finished 29-30 units. Consequently, A-LM students on the average were exposed to 525 new words, Holt classes to 650 new words, and Dale & Dale classes to 1400 (or 1500) new words. Now, if it is true that performance on the LA Form of the MLA Cooperative Reading Test is a function of vocabulary size, then we might predict that Dale & Dale students would do better than Holt and A-LM students. And this is precisely what happened. To confirm the importance of the vocabulary factor in this test, I analyzed each of the 50 items and found that the A-LM student who had mastered unit one through eleven would be able to answer 12 items correctly and perhaps get another two because of cognates.

---

*Paper read at the meeting of the NALLD, New York, December 28, 1968. Reprinted with permission of the authur and the National Association of Language Laboratory Directors.

He would have to guess on the remaining 38 items. The good Dale & Dale student, on the other hand, would be able to answer about 27 items correctly and would be forced to guess on the other 23. But the spread between the means of the "Audiolingual" classes and the "Traditional" classes is only about one and a half to four items: this might indicate that although the "Audiolingual" student is exposed to less vocabulary, he learns it better, and that the "Traditional" student cannot retain all that he is exposed to. This factor of vocabulary retention might well be the subject of further investigation.

The "Traditional" students also performed significantly better on the "critical sounds" section of the speaking test: here the student reads a passage aloud and is graded on his pronunciation of certain sounds. The "Traditional" students have had much practice in reading unfamiliar texts aloud whereas the "Traditional" /Audiolingual/ students only have read aloud material which they had already learned orally. Perhaps superior performance on this section is a function of the amount of practice.

Conclusion: In comparison to students using "Audiolingual" tests, first-year students using modified "Traditional" texts perform better on reading tests where size of vocabulary is a factor. They also perform better on tests of reading aloud.

2. How may the listening skill best be taught? The Pennsylvania project found no significant differences among teaching strategies or laboratory systems with respect to performance on the LA Listening Test of the MLA Cooperative Battery. All students, however, had the same number of weekly contact hours in foreign languages: five hours of classtime or four hours of classtime plus two half-hour lab periods. "Traditional" teachers were allowed to use the target language as much as they wished (except for grammar explanations), and it is quite possible that even the "Traditional" students heard the foreign language a good portion of the time. (This was not controlled by the project.) But, a significant difference on listening test scores was discovered when the students were grouped according to the text they used: in both German and French classes, the Holt students outperformed both the A-LM students and the "Traditional" students. The project report merely states that the two "Audiolingual" texts appear to be superficially similar. However, I have noted a difference which would explain the superior performance of the Holt students. The Holt series text is the only text among those utilized in the project which offers numerous recombined dialogs for each unit. The students are exposed to the structures and vocabulary of the lesson in a variety of situations. It is to be noted that all the recombined dialogs are printed in the student text. An area for further research would investigate relative effectiveness of such printed presentation versus a listening comprehension program available only on tape.

Conclusion: It would appear that if we wish to develop the skill of listening comprehension in our students, we must create materials which stress recombined dialogs and conversations.

3. What may we say about the future of the language laboratory at the secondary school level? We must admit that the laboratory as it has been generally utilized over the past several years has not contributed significantly to improving the students' "Audiolingual" skills. Does this mean we should scrap our laboratories and go back to the classroom tape recorder? Definitely not. But it does mean that we must find more effective ways to incorporate the laboratory into the foreign language classes. Perhaps drillwork is better conducted in the classroom, by the teacher or by tape. The new frontier of the language laboratory seems to open in two directions: the improvement of listening comprehension and the implementation of individualized instruction.

Listening Comprehension: As we noted earlier, frequent recombinations of known structures and vocabulary increase listening comprehension (as measured by the MLA Cooperative Listening Test.) Students need more listening practive. A variety of listening comprehension exercises (following maps, working out puzzles, playing Bingo) would probably also increase student motivation: winning a game is more fun than doing drills.

Individualized Instruction: In the language programs of the future, emphasis will fall on mastery. Students will master the basic core material of each lesson before advancing to the next lesson. For each lesson the teacher will have tapes at several difficulty levels: the faster students will practice understanding the foreign language at conversational and rapid conversational speed while the slower students will work with tapes on which speech is carefully enunciated. As language instruction moves toward more individualized programs, so will the laboratory play a more creative and more effective role in helping the student develop his language proficiency.

Conclusion: The "hardware" of the laboratory has undergone continual refinement over the past ten years, but the "software" has hardly changed. The challenge of the next decade will be the development of imaginative and more effective tape programs.

Exerpt from the minutes of the February 1969 meeting of COFLIC (Coordinators of Foreign Languages in Connecticut)

Pennsylvania Report

Ken Lester reported on an article in the "Newsfront" section of Education USA which oversimplied reporting of the results of the research study called "An Assessment of Three Foreign Language Teaching Strategies Utilizing Three Language Laboratory Systems." The "Newsfront" article reported that the study proved that the modern audiolingual method of teaching foreign languages is no more effective than the tranditional method.

Discussion resulted in the questioning of definitions of the "Functional Skills Method" (Audiolingual) and of the appropriateness of condemning this method rather than "this method as applied in this study."

Several weaknesses of sampling were pointed out which would make it unsound to apply the findings which were internally valid to the outside world of all foreign language study. Also, two of the tests given were not validated and no valid measure was taken of speaking ability.

Ken Lester read the list of "Recommendations" of the study, a much less sensational list than the summary of conclusions, the latter list being the one which the "Newsfront" article used. Ken agreed to have the recommendations and implications portions of the study duplicated and mailed to COFLIC members. It is these sections of the study which have significance for foreign language teaching in general.

Memorandum to COFLIC members - 2/13/69

From: Kenneth A. Lester, Foreign Language Consultant
Connecticut State Department of Education

Re: Final Report, Project No. 7-0133, USOE

A COMPARISON STUDY OF THE EFFECTIVENESS OF THE TRADITIONAL AND AUDIOLINGUAL APPROACHES TO FOREIGN LANGUAGE INSTRUCTION UTILIZING LABORATORY EQUIPMENT

Enclosed are the Summary (with conclusions), Implication and Recommendation sections of the above named report. I agreed to have these sections duplicated for you when we discussed this report at the COFLIC meeting February 7.

I have noted two more criticisms which you may find of interest. The Opinion Scale was not validated so the findings relative to attitude must be discounted.

Also, speaking and writing were not measured in this level two phase of the study. Tests on pronunciation and fluency, written by Rebecca Valette, turned out to be of questionable validity so no conclusions could be drawn about these two skills. (Please note that the first report, No. 5-0683, did measure these two skills and reported no significant difference between strategies. The report dealt with only level one.)

I encourage you to get the whole report of each of these studies if you expect to have to deal much with critics about the mis-reporting of what the studies "proved."

NATIONAL COUNCIL OF STATE SUPERVISORS OF FOREIGN LANGUAGES

February 26, 1969

Dear State FL Supervisor:

You should have received recently a copy of the final report on Project No. 7-0133, "A Comparison Study of the Effectiveness of the Traditional and Audiolingual Approaches to Foreign Language Instruction Utilizing Laboratory Equipment." The "Newsfront" page of Education USA recently circulated a news release on this project. The release stated, in part, that "The modern audiolingual method of teaching foreign languages is no more effective than the traditional method. That is the controversial conclusion of the Pennsylvania Foreign Language Project after repeating its experiment a second year to confirm its findings."

This oversimplification of the research findings is misleading and requires that the truth be pointed out by NCSSFL. Any research study must be read completely and with an open mind. All research of this type has some built-in weaknesses, since it cannot possibly be conducted under laboratory control conditions, and must be interpreted in the light of these deficiencies.

A careful examination of the research will show up several weaknesses of testing instruments, sampling techniques and operating definitions, as well as the standard difficulties of experimental control. NCSSFL suggests that you examine these weaknesses carefully in reviewing Phil Smith's research report.

The conclusions of the research, reported by Phil in the summary on vii and viii, based on the particular situation and subjects treated in this study, are not logically transferable in toto to the general field of foreign language instruction.

This is a sound piece of research, given the limitations of all experimentation of this type. It is misinterpretation which will trouble us. The investigators have considered the limitations in generalizing their conclusions. These generalizations, the only portion of the research which is honestly applicable to all of us, are reported in the "Implications" (page 112) and "Recommendations" (page 114) sections. We suggest that you read these sections carefully. Remember that even these represent only some more facts not of an entirely conclusive nature, and use them in your dealings with those who have jumped to unjustified conclusions after reading only a summary, out of context, of statements made in the research report which are of more sensational interest.

Kenneth A. Lester
President, NCSSFL

KAL:jfl

Enclosure

COMMONWEALTH OF PENNSYLVANIA
DEPARTMENT OF PUBLIC INSTRUCTION
West Chester State College
West Chester, Pennsylvania 19380

September 23, 1969

Mr. Kenneth A. Lester, Foreign Language Consultant
State of Connecticut
State Department of Education
Box 2219
Hartford, Connecticut 06115

Dear Mr. Lester:

I wish to thank you for the copies of your communications regarding our research to the National Council of State Supervisors of Foreign Languages and the Coordinators of Foreign Languages in Connecticut. These will be reproduced exactly as you wrote them and reproduced in our SUPPLEMENTARY REPORT.

I regret that you informed your colleagues that no valid measures were taken of speaking ability (COFLIC Minutes, paragraph 3) or of speaking and writing in Level II (COFLIC Memorandum, paragraph 3).

Both the MLA Cooperative Classroom Speaking and Writing Tests were given to a 10% random sample of all classes at Level I mid-year, the end of Level I, the end of Level II and for the Replication Study. (USOE 5-0683, pp. 63-69. USOE 7-0133, pp. 56 and 65-67. See Tables 23, 24 and 25). The data on Replicators was analyzed but notused due to the small number available in some treatments after the loss of a tape by a tester.

Simply because the Opinion Scale was not formally validated does not mean that it can be discounted (COFLIC Memo, paragraph 2). The instrument is heavily based upon the widely accepted work of Osgood. It correlates significantly with other indicies of student attitude and expectations (USOE 5-0683, Tables 107-110, pp. F-8 to F-11).

Even were it not externally valid, it has internal validity. What ever it measures, it can be assumed to measure the same factor for all students within the population, permitting comparisons such as we have made.

Some of our statistical analyses made with the Opinion Scale are open to question. These were made in input arrangement due to the limitations of our computer. We have been redoing some of these in a better manner with our newer computer but without very different results.

Thank you again for your courtesy in permitting us to use your materials.

Sincerely,

Philip D. Smith, Ph. D.
Project Coordinator

PDS/clk

APPENDIX A

SUMMARY OF STATISTICAL ANALYSES, FIRST AND SECOND YEARS OF THE STUDY

TABLE 1

TEACHING STRATEGIES

SUMMARY OF STATISTICAL ANALYSES AFTER ONE YEAR

Final Test	French (55 classes)		German (35 classes)	
Original: Analyses of Variance and Covariance	Prob.	Direction	Prob.	Direction
1. MLA Listening	NS		NS	
2. MLA Speaking*	NS		NS	
3. MLA Reading	NS		.03	TLM > AL
4. MLA Writing*	.003	TLM > AL	NS	
5. Coop. Reading	.001	TLM > AL	.001	TLM > AL
6. Coop. Vocabulary	.001	TLM > AL	.001	TLM > AL
7. Coop. Grammar	.001	TLM > AL	.001	TLM > AL
8. List. Discrimination	NS		NS	

	French (18 classes)		German (10 classes)	
Replication: Analyses of Covariance	Prob.	Direction	Prob.	Direction
1. MLA Listening	NS		NS	
2. MLA Reading	NS		NS	

---

* 10% random sample of each class

TABLE 2

TEACHING STRATEGIES

SUMMARY OF STATISTICAL ANALYSES AFTER TWO YEARS

Analyses of Covariance

Final Test	French II (24 Classes)		German II (25 classes)	
	Prob.	Direction	Prob.	Direction
1. MLA Listening	not sig.		not sig.	
2. List. Discrim.	not sig.		not sig.	
3. MLA Speaking*	not sig.		not sig.	
4. MLA Reading	.01	TLM>ALM	.05	TLM> ALM
5. MLA Writing*	not sig.		not sig.	

*10% random sample 21 classes French II, 21 classes German II.

TABLE 3

LANGUAGE LABORATORY SYSTEMS

SUMMARY OF STATISTICAL ANALYSES, BOTH YEARS

TR vs AA vs AR with Audiolingual Strategies

	FRENCH I (35 classes)	GERMAN I (24 classes)
Final Test	probability	probability
Original:		
1. MLA Listening	NS	NS
2. MLA Speaking*	NS	NS
3. MLA Reading	NS	NS
4. MLA Writing*	NS	NS
	FRENCH I (18 classes)	GERMAN I (10 classes)
Replication: (AA vs AR only)	probability	probability
1. MLA Listening	NS	NS
2. MLA Reading	NS	NS
3. List. Discrimination	NS	NS
	FRENCH II (24 classes)	GERMAN II (25 classes)
Follow up:	probability	probability
1. MLA Listening	NS	NS
2. MLA Speaking**	NS	NS
3. MLA Reading	NS	NS
4. List. Discrimination	NS	NS

---

* 10% random sample

** 10% rancom sample of 21 French II and 21 German II classes

TABLE 4

MEAN TEACHER MEASURES AND PROFICIENCY SCORES – TEACHERS WHO COMPLETED TWO YEARS OF INSTRUCTION

Training and Experience	French (N = 19)		German (N = 21)	
1. Graduate Semester hours:	36.42		44.48	
2. Yrs. teaching experience:	9.95		10.86	
3. Yrs. For. Lang. teaching:	6.84		7.52	
MLA Teacher Proficiency Tests:	Means	Nat'l %-ile	Means	Nat'l %-ile
4. Speak	37.74	50-55	41.81	60
5. Listen	71.00	60	88.52	65-70
6. Read	45.47	60	52.00	65-70
7. Write	44.42	55	57.00	65-70
8. Applied Linguistics	49.68	70-75	52.81	70-75
9. Culture	47.11	65	53.62	70-75
10. Professional Preparation	63.26	60	62.29	55

TABLE 5

RELATIONSHIP OF TEACHER FACTORS AND CLASS ACHIEVEMENT

Correlation Coefficients

Teacher Factors:	STUDENT MEASURES										
	1 Semester		1 Year				2 Years				
GERMAN, (N=21)	List.	Read	List.	Read	Speak	Write	List.	Read	Speak	Write	SOS
1. Type of Institution	.34	.08	.12	.08	.02	.17	.11	.06	.04	.31	.27
2. No. Grad. hours education	.17	-.05	.35	.22	.38	.38	.42	.01	.27	.50*	.02
3. Yrs. Teaching experience	.05	-.02	.07	.07	.29	.14	-.04	-.15	.23	.18	-.31
4. Yrs. F. L. Teaching	.00	-.04	-.00	.08	.35	.26	.04	.01	.25	.16	-.02
5. Self-est., Speaking	-.28	-.26	-.37	-.14	-.18	-.09	-.19	-.18	-.20	.19	.18
6. Self-est., Reading	-.32	-.01	-.15	.13	-.03	.14	-.04	-.19	-.12	.15	.13
7. Self-est., Writing	-.17	-.43*	-.15	.02	-.28	-.35	-.11	-.15	-.04	-.01	.01
FRENCH, (N=19)											
1. Type of Institution	-.35	-.31	-.01	-.03	.08	-.04	.02	-.14	-.18	-.24	.09
2. No. Grad. hours education	.08	.03	.11	.34	-.01	.04	-.03	.20	.26	-.06	-.01
3. Yrs. Teaching experience	-.27	.17	.10	.02	-.25	-.11	.11	.08	-.14	-.23	.36
4. Yrs. F. L. Teaching	-.18	-.14	.23	.07	-.23	-.04	.28	.15	-.04	-.20	.33
5. Self-est., Speaking	.06	-.40	-.22	-.10	.14	-.15	-.36	-.26	-.19	.03	-.20
6. Self-est., Reading	-.02	-.45*	-.25	-.34	.20	-.17	-.39	-.42	-.08	.05	-.18
7. Self-est., Writing	-.07	-.37	-.29	-.22	.25	-.03	-.47*	-.29	-.00	.16	-.31

* $p < .05$.

TABLE 6

RELATIONSHIPS OF TEACHER PROFICIENCY SCORES TO CLASS ACHIEVEMENT OVER A TWO YEAR PERIOD

Correlation Coefficients

Pre-Experimental MLA Teacher Proficiency Tests	STUDENT MEASURES												
	1 Semester			1 Year				2 Years					
	MLA List.	MLA Read	List. Disc.	MLA List.	MLA Read	MLA Speak	MLA Write	MLA List.	MLA Read	MLA Speak	MLA Write	List. Disc.	Opinion Scale
GERMAN (N = 21)													
1. Listen	-.21	-.33	.17	-.03	.12	.04	-.07	.11	.03	.25	.18	.20	.11
2. Speak	-.08	-.17	.18	-.03	-.03	-.00	-.05	.10	.09	.03	.14	.36	.16
3. Read	-.24	-.18	.21	-.12	.09	.11	.06	.02	.16	.16	.24	.34	.20
4. Write	-.10	-.19	.14	.02	.05	.26	.24	.21	.20	.19	.35	.31	.33
5. Linguistics	-.37	-.11	.22	-.23	-.16	.03	.00	-.10	.02	-.27	-.02	.38	.08
6. Culture	-.20	.04	.10	-.25	-.24	-.22	-.22	-.14	-.05	-.24	.05	.09	.27
7. Prof. Prep.	-.27	-.17	.20	-.27	.06	-.17	-.10	-.14	.20	-.11	.18	.26	.40
FRENCH (N = 19)													
1. Listen	.13	.04	.15	.54**	.30	-.19	-.05	.40	-.08	-.32	-.32	.28	.12
2. Speak	.24	.15	.59**	.55*	.32	.18	.27	.65**	.30	-.01	.07	.67**	-.24
3. Read	.06	-.06	.25	.51*	.15	-.28	-.12	.51*	.02	-.24	-.24	.29	.23
4. Write	.16	.12	.36	.40	.08	.02	.02	.46*	.06	-.06	-.05	.49*	-.13
5. Linguistics	.22	.27	.40	.47*	.17	-.20	.07	.39	.10	.10	-.01	.26	-.13
6. Culture	.09	.43	.70**	.27	.13	-.02	.13	.33	.08	-.13	-.20	.60**	-.12
7. Prof. Prep.	.33	.40	.38	.23	.12	.20	.10	.26	.07	.24	.00	.29	-.12

$*p < .05$, $**p < .01$.

APPENDIX B

THIRD AND FOURTH YEAR CLASSES AND SCHOOLS

THIRD YEAR TEACHERS AND SCHOOLS

FRENCH

	Teacher	School
105	Miss Joan Mesko	Nazareth Sr. H. S., Nazareth, Pa.
112	Mr. John Yoder	L. E. Dieruff H. S., Allentown, Pa.
122	Mr. William McDonald	Hampton Twp. H. S., Allison Park, Pa.
136	Mrs. Nancy Fisher	Wilson H. S., Reading, Pa.
151	Mrs. Joanna Clinchard	Lincoln H. S., Philadelphia, Pa.
153	Mrs. Donalda Costello	No. Allegheny H. S., Pittsburgh, Pa.
155	Mr. Richard Bond	Boyertown H. S., Boyertown, Pa.
162	Mrs. Geraldine Edsall	Mt. Penn H. S., Reading, Pa.
172	Mrs. Marguerite Fetterman	Cumberland Valley H. S., Mechanicsburg, Pa.
175	Mrs. Minerva Waldbaum	High School for Girls, Philadelphia, Pa.

GERMAN

	Teacher	School
202	Mr. Arthur Hollinger	Donegal H. S., Mt. Joy, Pa.
203	Mr. David Kruger	Annville-Cleona H. S., Annville, Pa.
204	Mrs. Ruth McGonigle	Nazareth H. S., Nazareth, Pa.
206	Mrs. Maria Schmid	Hatboro-Horsham H. S., Horsham, Pa.
213	Mr. Joseph Santer	Washington H. S., Philadelphia, Pa.
214	Mrs. Mally Shuster	Central H. S. Philadelphia, Pa.
243	Mr. Robert Reeser	Schuylkill Valley H. S., Leesport, Pa.
246	Mrs. Sophie Koshatka	High School for Girls, Philadelphia, Pa.
251	Miss Polly Clark	Palisades H. S., Kintersville, Pa.
252	Miss Marilyn Doebel	Bethel Park H. S., Bethel Park, Pa.
255	Mrs. Hedwig Voltz	Central Bucks H. S., Doylestown, Pa.
266	Miss Elsie Ewald	Olney H. S., Philadelphia, Pa.
272	Mr. Clark Schenck	Cumberland Valley H. S., Mechanicsburg, Pa.
283	Mr. Wilbert Wollenhaupt	Muhlenburg H. S., Lauraldale, Pa.

FOURTH YEAR TEACHERS AND CLASSES

FRENCH

	Teacher	School
162	Mrs. Geraldine Edsall	Mt. Penn H. S., Reading, Pa.
155	Mrs. Wilhelmine Lysinger	Boyertown H. S., Boyertown, Pa.
172	Mrs. Marguerite Fetterman	Cumberland Valley H. S., Mechanicsburg, Pa.
153	Mrs. Donalda Costello	No. Allegheny H. S., Pittsburgh, Pa.
151.	Mrs. Joanna Clinchard	Lincoln H. S., Philadelphia, Pa.
175	Mrs. Minerva Waldbaum	High School for Girls, Philadelphia, Pa.

GERMAN

	Teacher	School
203	Mr. David Kruger	Annville-Cleona H. S., Annville, Pa.
252	Miss Marilyn Doebel	Bethel Park H. S., Bethel Park, Pa.
202	Mr. Arthur Hollinger	Donegal H. S., Mr. Joy, Pa.
206	Mrs. Maria Schmid	Hatboro-Horsham H. S., Horsham, Pa.
283	Mr. Wilbert Wollenhaupt	Muhlenburg H. S., Laureldale, Pa.
251	Mrs. Ruth Gackenbach	Palisades H. S., Kintnersville, Pa.
266	Miss Elsie Ewald	Olney H. S., Philadelphia, Pa.
213	Mr. Joseph Santer	Washington H. S., Philadelphia, Pa.
214	Mrs. Mally Shuster	Central H. S., Philadelphia, Pa.
246	Mrs. Sophie Koshatka	High School for Girls, Philadelphia, Pa.
243	Mr. Robert Reeser	Schuylkill Valley H. S., Leesport, Pa.

APPENDIX C

PROJECT NORMS,

MLA COOPERATIVE CLASSROOM TEST (MA)

PROJECT NORMS

FRENCH, FORM MA, 6 SEMESTERS

Listening (N = 103)		Reading (N = 103)	
Raw Score	Percentile	Raw Score	Percentile
27-28	99	31-33	99
26	98	30	98
25	96	27	97
24	95	25	95
23	94	24	94
22	93	23	91
21	89	22	89
20	79	21	83
19	75	20	80
18	73	19	77
17	70	18	72
16	63	17	64
15	60	16	59
14	53	15	54
13	50	14	45
12	39	13	38
11	32	12	31
10	21	11	25
9	15	10	16
8	8	9	7
7	5	0-8	2
0-6	1	0-8	2

PROJECT NORMS

GERMAN, FORM MA, 6 SEMESTERS

Listening, N = 182

Raw Score	Percentile
29-31	99
28	98
27	97
26	96
25	93
24	90
23	89
22	85
21	81
20	79
19	75
18	69
17	66
16	60
15	56
14	48
13	39
12	27
11	19
10	14
9	9
8	6
7	8
6	2
0-5	1

Reading, N = 182

Raw Score	Percentile
29-31	99
28	98
26	97
24	96
23	95
22	94
21	93
20	90
19	87
18	84
17	78
16	74
15	68
14	58
13	50
12	42
11	31
10	23
9	16
8	10
7	5
6	3
0-5	1

APPENDIX D

LIST OF MANUFACTURERS OF LANGUAGE LABORATORY EQUIPMENT FOR SCHOOLS AND TREATMENTS

LANGUAGE LABORATORY MANUFACTURERS

Name of School	Treatment	Manufacturer of Laboratory
Methacton	AA	Lingua Trainer
Bartram	AA	Magneticon M.R.I. (T.R.W.)
Great Valley	AR	R.C.A.
Boyertown	AR	Magneticon
William Tennent	AA	Magneticon
Interboro	AR	Monitor
Conestoga	AA	Magneticon
Northern Pottstown	AA	Fleetwood Lingua-Center
Springford	AR	Magneticon
Plymouth-Whitemarsh	AR	Instructomatic
Beverly Hills	AR	Lingua Trainer
Allderdice	AA & AR	Magneticon
Elizabeth	AA	Magneticon
Peabody	AA	Magneticon
West Allegheny	AA	R.C.A.
Stowe	AA	Magneticon
North Allegheny	AR	Magneticon
Whitehall	AA	Magneticon-RCA Combination
Bethal Park	AA & AR	Magneticon M.R.I. (T.R.W.)
Mt. Lebanon	AA & AR	Magneticon M.R.I. (T.R.W.)
Churchill	AR	Magneticon
North Hills	AR	Magneticon
Fox Chapel	AA	Magneticon
Mt. Penn	AR	Rheem-Califone
Ephrata	AA	Fleetwood
Pen Argyl	AA	American Seating
Girls High, Phila.	AA	Magneticon
Lincoln, Phila.	AR	Magneticon
Central Bucks	AA & AR	Magneticon
Palisades	AR	Rheem-Califone
Easton	AA	Rheem-Califone
Scranton, Central	AR	R.C.A.
Emmaus	AR	Lingua Trainer (G.E.)

APPENDIX E

In Reply to the October 1969 *Modern Language Journal*

IN REPLY TO THE OCTOBER, 1969, MODERN LANGUAGE JOURNAL

A Talk by Dr. Philip D. Smith, Jr.

3rd Annual Meeting of the
American Council on the Teaching of Foreign Languages
New Orleans, November 28, 1969

When Henry Adams began his career a century ago with a critical analysis of the Captain John Smith-Pocahontas episode, he was advised that "it would attract as much attention, and probably break as much glass, as any stone that could be thrown...." This was not the intent of the Pennsylvania Foreign Language Projects--but it certainly seems to have been the case. It may be more appropriate to rename this portion of the program from "...on the Firing Line" to "Experts and Authors Meet the Firing Squad."

A decade ago the audiolingual revolution reached the American public schools. As an active participant, both as a state supervisor and a three-time NDEA Institute administrator, I planned many language laboratory installations, worked with many teachers, and was once told by a Harcourt, Brace and World representative that I was the best A-LM salesman west of the Rockies. I am proud to have been associated with a movement that restored life and vigor to foreign language education.

In 1967, I accepted the assignment of reporting the large-scale research studies in foreign language curriculum being completed by the Pennsylvania State Department of Education. In a sense the Modern Language Journal reviewers and I shared a common task, that of writing about a curriculum assessment planned, conducted, and designed by others. My task has been immeasurably easier than theirs by reason of two years full-support to complete the reports and by the access I enjoy to the files, the data collection, and the researchers, teachers, and students who participated.

I know, through personal experience, that the size alone of the Pennsylvania Projects--four thousand-two hundred students in one hundred and thirty-two classes representing an investment of three hundred and fifty thousand dollars and over a thousand pages of written materials--that size alone meant certain human oversights and errors in the conduct, the reporting, and in the interpretations and reactions to the findings. It was the hope of the Project Staff that our reports would elicit from the profession objective, scholarly, and thorough reviews. For this reason the research staff gave MLA-ACTFL and selected professionals six months to a year advanced notice of the forthcoming results in which to prepare the profession. We received no response. We were realistic enough to know that we could expect both responsible reactions and those who only saw in the Project a bogey-man of awesome proportions.

It is indeed unfortunate as we can always expect evaluation and judgment, whether we think we deserve them or not. The Pennsylvania

Studies may indeed be mild compared to how well foreign languages may fare if we are ever included in the forthcoming National Assessment.

Every member of the profession will be affected by the Pennsylvania Studies and their misinterpretation and misapplication. We cannot make them "Idiot Proof." I wish then, at the outset, to charge each member of ACTFL with the personal responsibility of becoming an objective and knowledgeable interpreter of the Pennsylvania Foreign Language Studies to the non-foreign language public.

Before proceeding, it is perhaps wise to admit that the Pennsylvania Foreign Language Research Studies were conceived and reported with certain biases and even naiveness. The study was undertaken as a reaction to the Keating Report and was an attempt by the Pennsylvania State Department of Education to support the already accepted state support of the audiolingual approach, the language laboratory and teacher certification by examination.

Pennsylvania was naive in that it fully expected, I believe, to vindicate the audiolingual approach and in that it believed that what was then considered a very carefully planned and conducted study would be accepted by objective professionals no matter what the outcome.

Since the text of the Pennsylvania reports are not yet available to the profession at large (ERIC processing is very slow), the Project finds itself in the incredible position of being reviewed but not widely read. The Modern Language Journal did not invite the Project to respond to the review nor did it accept the suggestion that the reviews be prefaced by a short description of the study to provide its readers with a better perspective.

At this point, may I especially commend Emma Birkmaeir, Dale Lange, and James Dodge for their care in pointing out what the Pennsylvania Studies do not prove. They do not prove anything. Few reviewers, with the exception of Valette and Carroll are interested in what our reports do say--and they do say a great deal. Specifically, I would like to suggest that the recent reviews published in the Modern Language Journal (October, 1969) often present a distorted view of the Pennsylvania Studies in that they suffer from (1) a narrow and insulated viewpoint; (2) overt hindsight; (3) personal interpretation; (4) inconsistency; and (5) obvious oversight. This is tragic, especially in that the Modern Language Journal attempts to be a responsible professional journal but will not protect its contributors nor its readers from obvious oversight, choosing to let errors stand as definitive statements on the research.

Since the MLJ reviews are to be "The last word" for many of the profession, I regret that in keeping the reviewers insulated from the Project Staff, the editor did my reviewing colleagues a serious disservice in that he permitted some to publish humanly preventable errors, oversights, and omissions that may now be to them a personal embarrassment.

Most reactors to the Pennsylvania studies view them much too narrowly, as that of a "tight" little experiment that somehow "got away from" the researchers. To the contrary, the studies were established purposely as a large scale assessment, (cf. the title, "An Assessment of ...") a curriculum innovation. Curriculum assessment, by definition, lags widespread acceptance. The study was planned on the advice of Campbell and Stanley that:

> ...experiments within schools must be conducted by the regular staff of the schools concerned...especially when findings are to be generalized to other classroom situations. (Gage, p. 191).

and Carroll:

> ...many questions concerning the education and training of foreign language students would have to be supported through experimental or longitudinal studies in which the effects of various types of learning experiences would be assessed by comparing pretest and posttest performances.... (Carroll, 1967, p. 207).

or, more recently, del Olmo:

> We should examine the list of characteristics of the audiolingual approach that have been isolated by Rivers (1964) and Valdman (1966), and show how these characteristics fare in the pragmatic atmosphere of the classroom (1968, p. 27).

and, lastly, Kerlinger (1965):

> ...research by no means needs to be limited to one variable at a time. It may even be said that it is wrong to so limit it, as Fisher has so strongly indicated.... (p. 229).

The Pennsylvania Studies represent an attempt to assess curriculum innovation in a "real life" situation--not as it might be, but as it is. Our reviewers, therefore, are too nearsighted and far, far removed from the realities of school district adoption when they suggest that every teacher should have been assigned the teaching strategy of his choice. Professor Hocking believes that our twice weekly use of the language laboratory was "sabotage." Not our use, but that of Pennsylvania secondary schools in 1964, again in 1968, and, according to the recent Clark-Austen survey (1969), of three-quarters of all secondary schools. The role of the state in establishing "exemplary" programs is a different matter. Hocking, in the opening paragraph of his review posits the effectiveness of the language laboratory in an idealized situation, without citations to supporting studies. Pennsylvania's assessment never pretended to exemplify the ideal teacher in the ideal situation with ideal students and its own laboratory maintenance specialist, but to determine if large scale foreign language innovations "suffered in translation." They did.

Insulation of the reviewers from contact with the Project staff led to serious errors which could have easily been corrected. Otto suspects that the teachers were not familiar with Teacher Manuals due to their cumbersome organization--a simple query would have told him that teachers only received pages pertinent to their individual assignment. Similarly, he raises questions concerning the content of the teacher training workshop, stating that it was a conference situation that "...did not provide exemplary models of effective teaching behaviors for each strategy." This is, regrettably, an assumption. More regrettable, it is not true. Good demonstration models were provided. The Project would have gladly provided a program of the meeting for him.

The admonition not to contact the Project staff debilitated many cogent comments. Due to the sheer size of the research reports, much important but secondary information had to be omitted. The answers to questions of Clark, for example, concerning comparisons of pre-experimental teacher factors or of covariance analyses without mid-year adjustments were his for the asking. (N.B. In this respect, may I compliment John Carroll for including in his forthcoming ACTFL review a day long visit to our offices with prepared questions, several direct inquiries, and the solicitation of additional computer analyses which we were more than happy to arrange for him. Although we can no longer do these operations, most are reported in our forthcoming SUPPLEMENTARY REPORT, and Project data is available on computer tape to interested professionals.) Our data is still being examined. My colleague, Emanuel Berger, for example, is currently examining some of the results of the Valette tests.

Lack of consistency among the reports is, I believe, an artifact of the isolation that may have been imposed by the Modern Language Journal. Valette points out that our analysis included scores in the "chance" range, a supplementary reanalysis with these scores deleted is criticized by Aleamoni and Spencer as invalid. Aleamoni and Spencer found nothing on laboratory maintenance but Hocking found enough, in his opinion, to invalidate the study. Otto questions the assumption of the MLA Proficiency Tests as predictors of student achievement while Clark rightly states they have not been validated in this respect--something the Project attempted to do and which will be reported in detail in the December Foreign Language Annals.

The Pennsylvania definitions and characteristics of teaching strategies are more concise than any others developed or published either before or after the research, including those for example, in the Chartain study that was accepted by the Modern Language Journal as viable research. It has been observed that Pennsylvania's criteria would have been hailed as precise and exemplary had the study only come out the way the profession expected.

It may be important to point out that several of the MLJ reviewers still hold the stereotyped view of the "traditional" teacher as the sort of mustachioed, black-hatted, "frito-bandito" that was common in the early sixties. Since then many have come to realize that old "Mrs. Traditional" was not, after all, inherently evil and that she did actually honestly try to teach a foreign language.

Our reviewers, like all of us, benefit from the illumination of hindsight. They should not, however, have overtly permitted it to intrude into 1969 criticisms made of a 1964 design and 1965-67 implementation. That they did is obvious in Otto's comments that we should have used the Pimsleur Tests, not released until mid-1967, --and reviewed research not yet reported until after the study was completed as is suggested by Aleamoni and Spencer. Professor Hocking continually cites my 1962 advice on language laboratory planning, but neglects to mention that it was a mimeograph hand-out available to the profession at large. Hocking also suggests that a minimum precaution on language laboratory facilities should have made reference to Language Laboratory Facilities by Hayes, dated 1968.

The reviewers, understandably, permitted personal interpretations to color their articles. Clark, for example, limited his summary of testing to the popular "four-skills" and omitted the unpopular but, according to Carroll, equally independent skills of reading by translation, vocabulary recognition, and explicit knowledge of grammar. I regret to see his otherwise fine review characterized by phrases like "may have" and "it is not difficult to imagine." Otto assumes that because the teachers took the MLA Proficiency Tests that they had an audiolingual bias when the teachers in reality had no option nor were they forewarned of the testing. Otto sees no relationship between required teacher proficiency levels and student achievement but Brooks, Freeman, Conant, and the Commonwealth of Pennsylvania did. Aleamoni and Spencer choose to view students as an unstable variable but overlook the fact that student scores were used. If scores are not viable our whole system of objective evaluation falls. Both Otto and Clark assume that teachers' comments about strategy assignments can be taken literally, not knowing that teachers had originally indicated two choices, permitting random assignment to make each teacher believe he got his preference. They did not consider the possibility that the teacher quoted was referring to another strategy with which he had little acquaintance (in fact the case if you know the speakers quoted).

Professor Valette has gone to great lengths to analyze the MLA Cooperative Classroom Tests and to state that, despite their 1963-64 reception as the long-awaited "audiolingual tests," they favor the "traditional" student. This may well be true. However, surely this must be balanced to some degree by the disadvantage of the "traditional" student taking a taped listening comprehension test for the first time especially in French classes who had never been exposed to a native speaker, or being subjected to reading and writing tests free from familiar English translation problems, or--for the first time--facing the traumatic experience of having to produce for a tape recorder actual foreign language speech.

The weight of the coin may be unevenly distributed, but surely it has two sides!

Dr. Valette dismisses consideration of any part of the study based on the MLA Cooperative Classroom Speaking Tests, stating they suffer from scorer unreliability. They may, eliminating in one fell swope both the esteemed F.S.I. rating scales and much foreign language research done in recent years. In the Pennsylvania Study, however, no more than two scorers ever worked with the Level I tests in either language. These randomly scored students within strategies and validated each other ($p < .01$ in German, $p < .05$ in French). For Level II, only one scorer worked in each language. The Speaking Test analyses should not, therefore, have been so easily discounted.

Otto suggests (p. 419) that "outdated versions" of the MLA Cooperative Classroom student tests were used. There has been only one version produced, and it is in widespread use in research, program evaluation, and college placement. The Project also defined the proportions of English and French or German that characterized the "traditional" class. Otto disagrees with this proportion and suggests that in a class where instruction is 3/4 English and 1/4 German, that English is not the predominate language.

I regret deeply that Professor Hocking is not present this morning.

Professor Hocking cites, at great length, Mr. Douglas Ward of Pittsburgh as an "inside source" to both comment on the research and to review the Hocking article before publication. I assume that Professor Hocking did this in good faith, I do believe he was ill-advised to accept Ward's contribution without, first, determing Mr. Ward's actual connection with the study; second, checking Ward's objectivity; and third; verifying Mr. Ward's statements.

I must admit that in three years of full-time work with the study, including several trips to Pittsburgh and visiting Project schools there, I have never met Mr. Ward. As a teacher at Taylor-Alderdyce High School, he, in no way, was in a direct position to be personally informed on what went on in most experimental and control classes.

Mr. Ward assisted the Project staff in the demonstration of language laboratory operation from August 22-25, 1965. Mr. Ward is best-remembered by the participating Project staff as a source of possible pre-experimental bias for a consistent and expressed negative attitude.

Mr. Ward was correct in informing Professor Hocking that six Pittsburgh teachers could not attend the pre-experimental workshop. His connection with the Project having terminated after only four days, he may not have even known that these teachers gave up the following two weekends, at the expense of their schools, for the necessary orientation.

Most of Ward's comments (p. 407) are dependent on hearsay evidence from a "project supervisor"--not one of the Project staff but presumably a school administrator--none of whom in Ward's area of the state was responsible for more than five of 104 classes which were visited much more often by Project observers than by local administrators or building principals. Mr. Ward also had no way in the world of knowing which of the classes he comments on were among the fourteen deleted from the statistical analyses. Ward states the laboratory was new to the teachers despite the pre-experimental requirement that a school could not participate without a laboratory the preceding school year. Ward's comment that the field representative never visited laboratory sessions can only be second or third-hand information and is refuted by dated observation reports and my personal observation. Ward's comments on laboratory quality can only accurately apply to schools of which he had a first-hand knowledge--three in Pittsburgh of the thirty-five used in the study.

In short, in a supposedly objective and unbiased professional critique, it first should have been observed whether or not a commentator really was in a position to know very much, and if he "had an axe to grind."

The Hocking review shows other instances of personal interpretation. Hocking takes the liberty of relating pre-experimental Project orientation and original language laboratory manufacturer orientation, beginning a statement with words from page 27 and concluding it with words from page 129 (pp. 405-406). Such liberties with context are not defensible.

Hocking states the Project was handicapped by late and hasty start despite two years of pre-planning including the Buch-Hayes Easton, Pennsylvania pilot study to which Hocking had earlier taken great exception. True, final approval was not made until the spring of 1965, a common case with federal funding, but literally months were spent in planning, discussions, and writing before first submission, professional readings, revisions, re-submission and contract negotiations. Even in the hey-day of the N.D.E.A. grant, amounts in hundreds of thousands of dollars involving over fifty different fiscal agencies were not obtained hastily.

Hocking's observation of the large turn-over in 1965 of the Pennsylvania State Supervisors is irrelevant. Continuity in this position has existed since 1963 and the state supervisors have always been kept informed but never actively involved in the conduct of the research. The results have been more directly disquieting to them than anyone in the profession.

The Aleamoni and Spencer article is most disturbing in its liberal use of personal interpretation, immediately assigning the study to **ex post facto** status despite the explicit paradigm of the research as Campbell-Stanley No. 10, the "Non-Equivalent Control Group" design. This should have at least have mentioned for the reader who is not familiar with the research reports even if the a priori ex post facto assignment by Aleamoni and Spencer is true, (which I and others believe it is not). The reviewers lament that (p. 424) no control group was used clearly overlooking the design paradigm which required none.

Aleamoni and Spencer are dissatisfied with our review of pertinent research (pp. 6-10), stating that it is not extensive enough, that it omitted summaries of previous research (i.e., "Gage" is cited as omitted but the review does include the correct reference to "Carroll") and the

Project is chided for omitting Pimsleur, dated two years after the Project was initiated (p. 423) or the 1967 Müller article on language laboratories.

The reviewers suggest a ladder approach. So did Pennsylvania. Based upon the Chicago, Colorado, New York and Easton Studies, Pennsylvania assumed that it was one of a succession not, as implied by Aleamoni and Spencer, the one-and-only.

The Student Opinion Scale (semantic differential) used is taken to task as ignoring Osgood when it was largely taken from his work. The scale's evaluation as meaningless since no relationship to an absolute was established (p. 424) is a complete misinterpretation of its intent to provide relative ratings among teaching strategies.

Aleamoni and Spencer imply (p. 425) that the Pennsylvania State Department of Education, in order to establish a "control group," should have tested classes in schools which had already indicated an unwillingness to permit this. No state would find this possible.

They also suggest that it was not necessary to drop students with incomplete data. The analyses of covariance could not legitimately accept a large number of generally low-aptitude, low intelligence scores as part of a covariate for subsequent achievement by more apt or able students.

Spencer and Aleamoni overlook many points obvious in the text of the reports: (2) they state that no information on the relationships among the dependent variables is a lack of control when correlations are in fact reported; (b) "No research date" on teacher certification by examination overlooks pages 106-123 inclusively; (c) our recommendation that, since language laboratory recording during the class period seemed to have no effect on achievement, labs should still have recorders for testing purposes was dismissed with the irrelevant comment that "No data collected related to the content of language laboratory tapes." This also ignores the fact that the text of tapes is printed in the books; (d) "No data was objectively nor systematically collected" on laboratory maintenance overlooks pages 128-129, more than obvious to Hocking and the stacks of maintenance reports in our files; (e) "No data is presented" on listening tests as predictors of student achievement ignores this factor as significant in twenty multiple regression equations, Tables 24, 25, and 26.

Several misinterpretations are obvious, including the reading of variation as variance (p. 427); the interpretation of my phrase "implication for generalization" as "implication in favor of generalization" when the context signifies the exact opposite (p. 427); the contention that the study contradicts its own data when it states that "curriculum innovations...have been widespread" but more superficial than the profession had hoped simply cannot be true. Surely, this is what the reports do say--that the audiolingual approach and the language laboratory did not have the effect we had expected.

Lastly, it is not clear at all why responsible educators and research specialists, consigning the Pennsylvania studies to the proverbial "roundfile," cannot accept the concluding statement that "more study is needed to advance knowledge of the second language learning process in the realistic setting of the public school."

In conclusion, my personal reaction to the reviews ranges from admiration to tears. Mostly tears because I feel that the reviews could have been both better done and more constructive.

The Pennsylvania studies make no pretense at being either definitive or flawless. May I urge each member of the profession to obtain and study the complete text since, as Valette responsibly points out, there are some meaningful implications in the reports for the profession.

We are not yet doing as good a job as we told our clientel we could; we have never given our technology an even break in keeping it within the class period; American secondary students still do not feel a functional command of a foreign language is important.

Our clientel, students and parents, and our colleagues in other classrooms and school offices do not read the Modern Language Journal. They do read the simplified and overstated summaries printed in the professional and public press. Only each of you here today can effectively interpret both the studies, the reviews, and the simplistic versions for those "outside the pale."

Sure, there is controversy. This is good. As Benjamin Harris stated, "In a changing society, a state of peaceful calm without friction is likely to mean either that nothing is going on or that what is going on is so far removed from the significant events of life that it doesn't matter." To me, foreign language education does matter. Thank you.

OE 6000 (REV. 9-66)

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

ERIC REPORT RESUME

(TOP)

ERIC ACCESSION NO.

CLEARINGHOUSE ACCESSION NUMBER

RESUME DATE	P.A.	T.A.
9 25 69		

IS DOCUMENT COPYRIGHTED? YES ☐ NO ☒

ERIC REPRODUCTION RELEASE? YES ☒ NO ☐

TITLE

Supplementary Report of USOE Project 7-0133

PERSONAL AUTHOR(S)

Smith, Philip D. Jr.

INSTITUTION (SOURCE)

Bureau of Research, Dept. of Education, Comm. of Penna.

SOURCE CODE

REPORT/SERIES NO.

OTHER SOURCE

Bureau of Research, West Chester State College

SOURCE CODE

OTHER REPORT NO.

OTHER SOURCE

SOURCE CODE

OTHER REPORT NO.

PUB'L. DATE 9 - -69

CONTRACT/GRANT NUMBER OEC-1-7-070133-0445

PAGINATION, ETC.

133 pp., Appendixes

RETRIEVAL TERMS

a. Foreign Language
b. French
c. German
d. Teaching Method
e. Language Laboratory
f. Teacher Preparation

IDENTIFIERS

Project 7-0133, Pennsylvania Foreign Language Project

ABSTRACT

USOE Projects 5-0683 and 7-0133 were continued into Levels III and IV. The conclusions support earlier findings that students learning foreign language through a cognitive approach continue to achieve significantly better than audiolingual students.

Comments of several noted foreign language educators question various aspects of the study. Reanalysis of original data continue to support earlier conclusions.